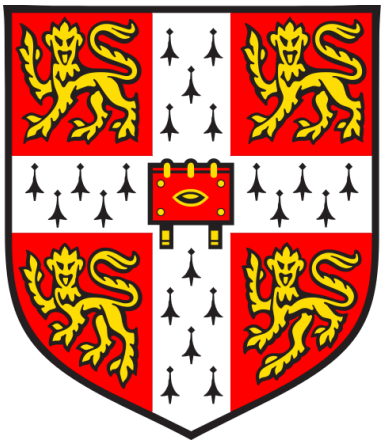
Structural and biophysical analysis of Human DNA repair protein CtIP

Christopher Robert Morton

Robinson College

University of Cambridge



This dissertation is submitted for the degree of

Doctor of Philosophy

December 2018

Declaration

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the main text. This work was carried out under the supervision of Professor Luca Pellegrini in the Department of Biochemistry of the University of Cambridge. It is not the same as any I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution. It does not exceed the prescribed word limit for the Degree Committee for Biology.

Cambridge, December 2018

Christopher Robert Morton

Abstract

The integrity of our genome is constantly threatened by endogenous and exogenous sources of DNA damage. The successful repair of DNA lesions is necessary for cellular survival and prevention of disease such as cancer. Therefore, complex molecular pathways have evolved to allow for accurate and timely DNA repair. It is now well-established that mammalian CtIP (CtBP-Interacting Protein) has an important function in DNA Double-Strand Break repair, by promoting the resection of DNA ends in preparation for homologous recombination (HR) and micro-homology mediated end joining (MMEJ). However, the biochemical basis of CtIP's critical role in DNA double-strand break repair remains poorly understood.

The aim of this work is to improve our understanding of CtIP's function in maintaining genomic stability. In my thesis I describe the structural and biophysical characterisation of human CtIP and its two evolutionarily conserved N- and C-domains that are critical for CtIP function. Sufficient quantities of recombinant CtIP were expressed and purified for *in vitro* analysis. A combination of biophysical approaches, including analytical ultra-centrifugation, size-exclusion chromatography and cryo EM, were used to probe the oligomeric nature of CtIP. All techniques showed the presence of high molecular weight oligomeric CtIP species.

The N-terminal region of CtIP has been previously shown to be essential for its role in DNA end resection, HR and MMEJ. Using X-ray crystallography in combination with small-angle X-ray scattering I demonstrate that an extended region of the CtIP N-terminus spanning amino acids 31 to 136 forms an elongated parallel helical dimer with coiled-coil segments flanking a central zinc-binding motif. In combination with previous crystallographic analysis of the N-terminal tetramerisation motif of CtIP, my new structural information allows modelling the overall architecture of the CtIP NTD.

I have explored the ability of CtIP to bind DNA, via electrophoretic mobility shift assays and fluorescence polarisation, to determine the affinity and minimum binding site of its interaction with double-strand DNA. Using alanine scanning mutagenesis, I have investigated the zinc-binding ability of putative metal-coordination residues within the C-terminal domain. Furthermore, I provide evidence for the presence of a coiled-coil region N-terminally juxtaposed to the CtIP-CTD, which might mediate its dimerization.

My findings advance our current knowledge of the biochemical and structural properties of human CtIP. They represent a useful contribution to our current efforts to improve our limited understanding of the molecular mechanisms presiding over the critical DNA-end resection step of HR.

Acknowledgements

I would like to take this moment to thank everyone who has helped me throughout my journey to reach this point in my life.

I would like to express my upmost gratitude to Professor Luca Pellegrini for enabling me to carry out the research in his laboratory, providing support and guidance throughout my PhD. Thank you for encouraging my scientific thinking to steer projects through the use of many interesting techniques to develop my scientific understanding. Thank you to Dr Neil Rzechorzek for your patience, supervision, teaching, time and understanding throughout my PhD. You have truly been indispensable both in my scientific and personal development. Thank you to Dr Mairi Kilkenny for all your support, encouragement, and teaching. Thank you to Dr Joseph Maman for teaching me valuable techniques and detailed scientific discussions. Further I am thankful to members of the Pellegrini, Robinson, Leadlay, and Hyvönen laboratories both past and present during my time in the department for scientific discussions and memorably times outside of the laboratory.

This project was enabled by a Wellcome Trust Senior investigator award to Professor Luca Pellegrini. This award funded my employment as a Research Assistant and laboratory expenses during my PhD. I would like to thank, Dr Paul Brear and Simon Quick for assistance using the X-ray crystallography facility, Dr Dimitri Chirgadze, and Dr Jamie Blaza for the use of the Cryo-EM facility, Dr Katherine Stott for help using the biophysics facility and the members of the DNA sequencing facility. I would like to especially thank Peter Sharratt and Dr Len Packman for their excellent mass-spectrometry service which will be greatly missed. Further to this, thank you to members of Diamond Light Source who provided support with X-ray data collection, Dr Robert Rambo and Dr Nathan Cowieson for assistance using the B21 SAXS instrument, finally Dr Pierre Aller, Dr Halina Mikolajek and Dr Juan Sanchez-Weatherby for assistance with HC1 dehydration experiments.

To my mum, Elaine Morton, dad, Christopher Morton and my partner, Anna Reva. This thesis is for you. Thank you. Thank you for all your unconditional loving support, encouragement, and making the dark days of science bright. Making the days when things go wrong bearable and joining the euphoric delight of a discovery. I would like to thank my sister, Dr Sarah Morton for sparking my initial desire to learn and push the boundaries of science, alongside my brother, Alan Morton, Peter Kimber and my grandparents for always being there for me with support. Finally, I would like to thank Anna's parents Tanya and Sergejs Reva and my friends, Neil Nathwani, Dr Grace Adams, Dr Luke Thompson, Aneeka Ul-Haq and Hannah Roberts for being there for me throughout my PhD, as well as customers of my laser engraving hobby for helping to keep me sane!

Table of Contents

List of figures

Chapter 1 figures, Introduction

Chapter 2 figures, Methods

Chapter 3 figures, CtIP N-terminal Domain

Chapter 4 figures, CtIP C-terminal Domain

Chapter 5, CtIP Full-Length

Chapter 6, Discussion

Appendix.

List of abbreviations

Abbreviation	Expansion
53BP1	Tumour suppressor p53-binding protein 1
9-1-1	Rad9–Hus1–Rad1
AAA	Amino Acid Analysis
ABC	ATP-binding cassette
ADP	Adenosine diphosphate
AND-1	Acidic Nucleoplasmic DNA-binding protein 1
AP	Apurinic/apyrimidinic
APC/C	Anaphase-promoting complex / cyclostome
APE1	Apurinic/apyrimidinic endonuclease 1
APS	Ammonium persulphate
AT	Ataxia Telangiectasia
ATM	Ataxia Telangiectasia Mutated
ATP	Adenosine triphosphate
ATR	ATM and Rad3 related
ATRIP	ATR interacting protein
AUC	Analytical Ultra-Centrifugation
BACH1	BRCA1-associated C-terminal helicase
BARD1	BRCA1-associated RING Domain 1
BER	Base Excision Repair
BLM	Bloom syndrome protein
BRCA1/2	Breast cancer type ½
BRCT	BRCA1 C terminus
BSA	Bovine Serum Albumin
CAK	CDK activating kinase
Caspase	Cysteine-aspartic proteases
CBX4	SUMO-protein ligase 4
Cdc25	Cell division cycle 25
CDK	Cyclin Dependent Kinase
cDNA	Complementary deoxyribonucleic acid
ChIP	Chromatin immunoprecipitation
Chk	Checkpoint Kinase
CMG	Cdc45-MCM-GINS complex
CRISPR-Cas9	Clustered regularly interspaced short palindromic repeats – caspase 9
CSB	Cockayne Syndrome Group B
CtBP	C-terminal binding protein of adenovirus E1A
CTD	C-terminal domain (of CtIP)
CtIP	CtBP interacting protein
DDR	DNA Damage Response
DE3	Phage genome DE3
D-loop	Displacement loop
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid

DNA-PKcs	DNA-dependent Protein Kinase, catalytic subunit
DSB	Double Strand Break
DSBR	DNA double strand-break repair
dsDNA	Double stranded DNA
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
EMSA	Electrophoretic mobility shift assay
EPR	Enzymatic Photo-Reactivation
ESRF	European Synchrotron Radiation Facility
EXO1	Exonuclease 1
FA	Fanconi Anemia
FANC	FA complementation group I
FANCD2	Fanconi Anemia complementation group D2
FANCM	Fanconi Anemia complementation group M
FEN1	Flap endonuclease 1
FHA	Forkhead-associated domain
FP	Fluorescence polarisation
FRET	Fluorescence resonance energy transfer
G1 / G2	Gap phase 1 / Gap phase 2
GADD45	Growth arrest and DNA damage-inducible gene 45
GG-NER	Global genome nucleotide excision repair
H2AX	H2A histone family member X
HDR	Homology-directed repair
HEK293F	Human Embryonic Kidney 293-FreeStyle™ cell line
HJ	Holliday junction
HNPCC	Hereditary nonpolyposis colorectal cancer
HR	Homologous Recombination
HSP70	Heat Shock Protein 70
ICL	Interstrand crosslink
IPTG	Isopropyl β-D-1-thiogalactopyranoside
IR	Ionising Radiation
KLHL15	Kelch-like protein 15
Ku	Ku 70/ Ku 80
LB	Lysogeny broth, Luria broth
LMO4	LIM domain only 4
MALDI-TOF MS	Matrix-Assisted Laser Desorption/Ionisation – Time of Flight Mass Spectrometry
MALS	Multi-Angle (laser) Light Scattering
MAPK	Mitogen Activated Protein Kinase
MBP	Maltose-binding protein
MDC1	Mediator of DNA damage checkpoint protein 1
MGMT	O-6-methylguanine-DNA methyltransferase
MIK1	Mitosis inhibitor protein kinase
MMEJ	Microhomology-mediated end joining
MMR	DNA mismatch repair
MPD	2-Methyl-2,4-pentanediol
M-phase	Mitosis phase

MR	Molecular Replacement
MRN	Mre11 Rad50 Nbs1
mRNA	Messenger ribonucleic acid
MRX	Mre11-Rad50-Xrs2
MTOR	Mammalian Target Of Rapamycin
MWCO	Molecular Weight Cut Off
NBS1	Nijmegen Breakage Syndrome 1
NEB	New England Biolabs
NER	Nucleotide excision repair
NHEJ	Non-homologous End-Joining
NTA	Nitrilotriacetic acid
NTD	N-terminal domain (of CtIP)
OD_{600}	Optical density at 600 nm
PAGE	Polyacrylamide gel electrophoresis
PALB2	Partner and localizer of BRCA2
PAR	4-(2-pyridylazo)-resorcinol
PARP	Poly (ADP-ribose) polymerase
PAXX	Paralog of XLCC4 and XLF
PBS	Phosphate-buffered saline
PCNA	Proliferating cell nuclear antigen
PCR	Polymerase chain reaction
PEG	Poly-ethylene glycol
PEI	Polyethylenimine
PES	Polyethersulfone
PIC1/2/3	Pre-incision complex 1, 2 and 3
PIKK	Phosphatidylinositol 3-kinase-related kinase
PIN1	Peptidyl-polyl cis-trans isomerase
PKM2	Pyruvate kinase isozyme M2
PLK1	Polo-like kinase 1
PML	Promyelocytic leukaemia
PNKP	polynucleotide kinase/phosphatase
Pol	Polymerase
PSI	Pounds per square inch
RB	Retinoblastoma
RBBP8	Retinoblastoma-binding protein 8 (Or CtIP)
RBP-JK	Recombining binding protein suppressor of hairless homolog
RCF	Relative centrifugal force
RECQL	ATP-dependent DNA helicase Q1
RFC	Replication Factor C
RI	Refractive index
RIF1	Replication timing regulatory factor 1
RIM	Alternate name for CtIP
RMSD	Root mean square deviation
RNA	Ribonucleic acid
RNF8	Ring Finger Protein 8
ROS	Reactive oxygen species
RPA	Replication Protein A
RPM	Revolutions per minute
Sae2	Sporulation in the Absence of spo 11

SAXS	Small Angle X-ray Scattering
SDS	Sodium dodecyl sulphate
SDSA	Synthesis-dependent strand annealing
SEC	Size Exclusion Chromatography
SeMet	Selenomethionine
SIAH1	Seven in absentia homolog 1
siRNA	Small interfering ribonucleic acid
SMC	Structural maintenance of chromosomes
SOC	Super Optimal broth with Catabolite repression
S-phase	Synthesis phase
SSB	Single Strand Break
Strep-Tactin	Strep-Tactin™ Superflow®
SUMO	Small ubiquitin-like modifier
TB	Tris Borate
TBE	Tris Borate EDTA
TBP	TATA-binding protein
TCEP	Tris(2-carboxyethyl) phosphine
TC-NER	Transcription-coupled nucleotide excision repair
TDP2	Tyrosyl-DNA phosphodiesterase 1
TE	Tris-EDTA
TEM	Transmission electron microscope
TEMED	Tetramethylethylenediamine
TEV	Tobacco etch virus
TFIIB	Transcription initiation facto IIB
TFIIH	Transcription factor II H
TopBP1	topoisomerase binding protein1
U2oS	Human bone osteosarcoma cells
UV	Ultraviolet
UVP	Ultra-Violet Products
WRN	Werner syndrome ATP-dependent helicase
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside
XLF	X-ray repair cross-complementing protein 4-like factor
XPA	Xerodema pigmentosum complementation group A
XP	Xeroderma pigmentosum
XPC	Xerodema pigmentosum complementation group C
XPG	Xerodema pigmentosum complementation group G
XRCC1	X-ray repair cross-complementing protein 1
XRCC4	X-ray repair cross-complementing protein 4
γ-H2AX	Phosphorylated H2AX

Chapter 1, Introduction.

1.1 DNA and genome stability

Nucleic acids are the molecular carriers of genetic information. All life forms depend on nucleic acids, as the information encoded by their sequence of bases: guanine, adenine, cytosine and thymine, is fundamental to an organism's existence. Not only do they encode the organism's hereditary information, but they encode all the functional and structural characteristics of the cell. Therefore, eukaryotic cells devote a vast amount of resources to maintain the integrity of their genomes against the ever-present threat of genotoxic agents and processes.

Deoxyribonucleic acid (DNA) is a highly stable biological macromolecule, which makes it a well suited as chemical medium for storage of genetic information. The closely related polymer of ribonucleic acid (RNA) is more susceptible to decomposition. This vulnerability can be found where high mutation rates are beneficial as, for example, in the small genomes of RNA viruses. Both DNA and RNA are constructed with basic building blocks, the nucleotides. The human genome consists of approximately 3,200,000,000 nucleotides of chromosomal DNA, which undergo an estimate of 100,000 spontaneous chemical reactions to form DNA lesions per cell per day (Hoeijmakers, 2009; Saul and Ames, 1986).

1.2 Causes, types and risks of DNA damage

Unwanted damage to the chemical structure of nucleic acids have been a constant threat to faithful transmission of genomic information since the very early life forms developed on earth (Kitadai and Maruyama, 2018). A DNA lesion is an alteration to the chemical structure of the DNA, affecting a base or a nucleotide, or causing a single- or double- strand break (SSB or DSB) in the phosphodiester backbone. These lesions or 'DNA damage' can have either endogenous or exogenous sources.

Endogenous DNA damage can occur in many ways such as through spontaneous hydrolysis of the purine and pyrimidine bases, replication errors inserting incorrect bases, oxidative metabolism and non-enzymatic methylation. Common examples of DNA damage include deamination of cytosine to give uracil that, if left un-repaired, would pair with adenine causing a C to T transition, and oxidative metabolism that produces reactive oxygen species (ROS) which can yield oxidized bases such as 8-oxoguanine and damage to the phosphodiester backbone (Hoeijmakers, 2009; Jackson and Bartek, 2009; Norbury and Hickson, 2001; Valko et al., 2006).

Exogenous DNA damage can be produced by agents such as genotoxic chemicals or high-energy radiation. Solar ultraviolet (UV) radiation is one of the most prevalent forms of environmental DNA damaging agents. The UV radiation spectrum ranges from 190-400nm, with the more dangerous UV radiation (<280nm) absorbed by the ozone layer; however UV A (315-400nm) and UV-B (280-315nm) radiation can produce 10,000 photoproducts such as cyclo-butane pyrimidine dimers per cell per day (Hoeijmakers, 2009). In addition, UV radiation can indirectly cause DNA damage through the production of ROS within cells (Jackson and Bartek, 2009). Ionizing radiation (IR), such as X-rays, generates various forms of DNA damage, for example DNA-protein adducts, base damage, SSB and DSB. Among these damage events a DSB is the most detrimental lesion due to the creation of unprotected DNA ends. Chemical agents causing exogenous DNA damage can be both the cause and treatment of cancers through mutation of coding DNA. For example, polycyclic aromatic hydrocarbons and nitrosamine 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone found in tobacco smoke form DNA-adducts that predispose to lung cancer (Hecht, 1999). However, chemotherapy agents, such as anthracyclines can cause apoptosis through excessive DNA damage to treat lung cancers (Cheung-Ong et al., 2013; Ciccia and Elledge, 2011).

Cells have evolved efficient mechanisms to repair DNA and maintain genomic integrity. Failure of these mechanisms can lead to detrimental consequences such as various forms of cancer as well as developmental and neurodegenerative disease. Furthermore, predisposition to DNA damage can be inherited via one of the rare genome stability syndromes, such as Nijmegen breakage syndrome, ataxia telangiectasia, Fanconi anemia, Bloom Syndrome, and Werner syndrome (Knoch et al., 2012).

These genetic conditions have proven extremely useful model systems to elucidate the molecular mechanisms that preserve genomic integrity in the cell.

1.3 The DNA damage response

In response to DNA damage, multiple repair mechanisms have evolved that help maintain genome integrity and assist in the faithful replication of DNA. A vast network of protein-protein interactions and modifications has emerged which sense DNA damage and transduce this information appropriately, resulting in the activation of a complex series of lesion type-dependent pathways that are collectively known as the DNA Damage Response (DDR). This cellular DDR can lead to changes in gene transcription, upregulation of the necessary proteins for repair, cell cycle arrest, or as a last resort senescence or apoptosis to ensure damage is not inherited (Ciccia and Elledge, 2011; Jackson and Bartek, 2009; Polo and Jackson, 2011).

1.3.1 The eukaryotic cell cycle

Alongside day to day tasks of routine metabolism the cell must duplicate its genome at an appropriate time during its cell cycle and prior to cell division. This process includes a multitude of pathways necessary for successful replication of the genome and division into two daughter cells (Tyson and Novak, 2008). The cell cycle can be divided into four distinctive phases: gap phase G1, synthesis phase S, gap phase G2 and Mitosis M phase. Replication of the genome occurs in S phase with cell division occurring in M phase. The two gap phases G1 and G2 allow sufficient time for preparation of the next phase in the cell cycle. Cells require a period of growth to ensure they are metabolically active before they can pass the G1-S phase checkpoint (also known as restriction point) and start DNA synthesis. Throughout S-phase cells are replicating their DNA so that each chromosome gives rise to two sister chromatids that are held together by cohesion factors. Following S-phase, a further gap phase G2 ensures the cells have enough time to complete DNA synthesis and repair of each chromosome and are ready for cell division. As all sister chromosomes are available in G2 Homologous Recombination can now take place to repair complex breaks. Once cells enter M phase chromosomes are bound via kinetochores to spindles at opposite poles and become aligned along the metaphase plate. Following correct segregation new nuclei begin to form at opposite sides of the original cell and cells can undergo cytokinesis to form two daughter cells, each with an exact copy of the parental cell's DNA.

This entire process is tightly controlled by restriction checkpoints at various stages of the cell cycle. The G1-S phase checkpoint is arguably the most critical as it separates the post-M from the pre-S

interval. Passing this checkpoint is necessary to continue to replicate. At this point cells can either commit to a further round of replication entering S-phase, or can withdraw from the cell cycle into G0 phase whereby they do not replicate their DNA and become senescent (Zetterberg et al., 1995). Cell-cycle checkpoints are highly conserved among eukaryotes and depend on the activity of cyclin-dependent protein kinases (CDKs) in cooperation with their binding partner cyclin. The abundance of CDKs remains constant throughout the cell cycle. However, two main processes have evolved for regulation of CDK activity: binding to different cyclin partners to target different protein substrates and post-translational phosphorylation to control activity. The levels of cyclin expression are dependent on the cell stage, with their rate of transcription and degradation controlling the abundance. This process allows activation of different cyclin CDK complexes to mediate the cell cycle stage. For example, the Cyclin E CDK2 complex is activated for entering S-phase, however this is inhibited during M-phase with Cyclin B CDK1 activated to control mitosis (John et al., 2001; Murray, 2004). Further counteracting proteins are involved in controlling the activity of CDKs; Wee1-like kinases cause inhibitory phosphorylation on a CDK's activation loop, while Cdc25 can dephosphorylate and reverse the inhibitory effect. Further control is through the CDK-activating kinase (CAK), which on phosphorylation allows full CDK activity (Coleman and Dunphy, 1994; Morgan, 1995).

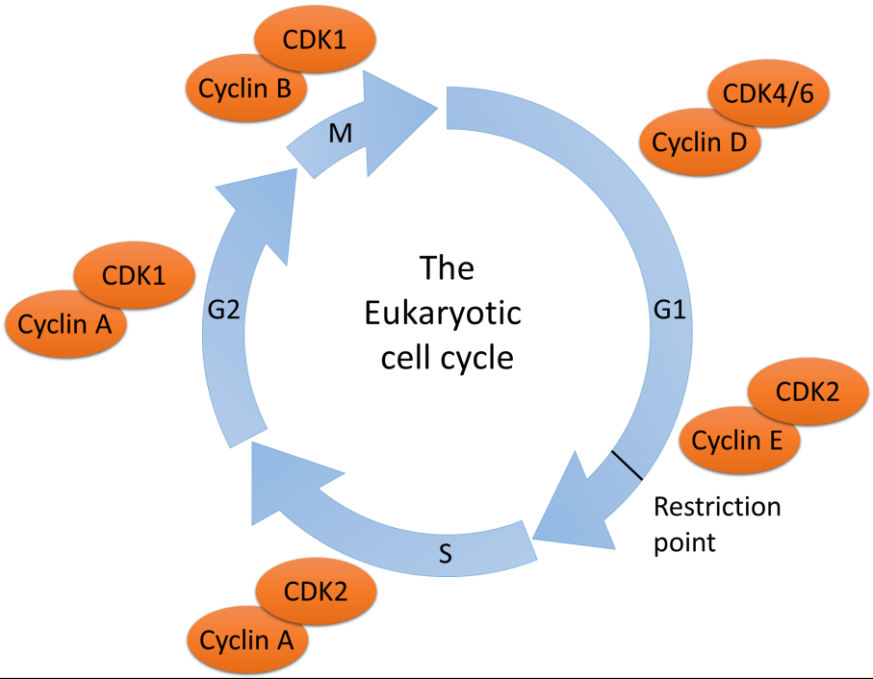


Figure 1.3.1: **The Eukaryotic cell cycle.** The Eukaryotic cell cycle contains four phases, G1, S, G2 and M. Progression through the cell cycle is governed by the checkpoint controls in normal cells. Cells can exit the cell cycle into G0 or by becoming senescent.

In yeast *S. cerevisiae* and *S. pombe*, where a lot of the original CDK research was performed, only one CDK is present to control the cell cycle checkpoints. In contrast, humans have at least 10 members of the CDK family: CDK1, 2, 4 and 6 are known to be involved in cell cycle control with 4 different cyclins

A, B, D and E (figure 1.3.1) (Cooper, 2000; Meyerson et al., 1992). Further control of CDK activity comes from proteins which inhibit cyclin-CDK complexes and in some cases activate others (Besson et al., 2008). The cyclin-CDK complexes are also a target for cancer treatment as checkpoint inhibition can lead to apoptosis, which will be discussed later in this section.

1.3.2 DNA-damage induction of cell cycle checkpoints

DNA damage can elicit a rapid and sensitive response to slow or even arrest cell cycle progression. Checkpoints are therefore a constant surveillance mechanism for stress and DNA damage, as incorrect decision making and repair can potentially lead to uncontrolled cell proliferation and cell death, especially in the context of the most genotoxic lesions, DSB. While many lesion-specific responses for DNA repair exist, they all activate common checkpoint pathways with the goal to maintain CDK activation until the lesion is removed. This response can be detailed by four components: the sensors (such as ATM and ATR), mediators which amplify the detection (such as BRCA1, 53BP1 and MDC1), the transducers (such as Chk1 and Chk2) and the effectors (such as p53 and Cdc25) (Barnum and O'Connell, 2014; Sancar et al., 2004). The simple checkpoint activation pathway leads from sensors detecting DNA damage and abnormalities that need repairing through mediators to relay the signal to the transducers. Transducers can then regulate the effectors that directly control and inhibit the cell cycle progression to allow sufficient time for repair of the DNA lesion. These pathway stages, however, can be attributed to a single protein acting at several levels: for example ATM can act as a sensor, mediator and transducer of DNA damage with multiple roles within a checkpoint response (Barnum and O'Connell, 2014; Sancar et al., 2004).

1.3.3 ATM and ATR activation by DNA damage

Protein complexes exist to transduce alterations in the DNA in the initial stages of sensing DNA damage (Harper and Elledge, 2007). Of these, Ataxia-Telangiectasia Mutated (ATM) and ATM and Rad3 Related (ATR) are two of the most studied early sensors/mediators of the DNA damage. They belong to the family of phosphatidylinositol-3 kinase-related serine/threonine kinases (PIKK) which also includes mammalian target of rapamycin (MTOR), suppressor of morphogenesis in genitalia (SMG1) and DNA-dependent protein kinase catalytic subunit (DNA-PKcs). Functions of the PIKK family span a wide range of pathways including cell survival and metabolism (Weber and Ryan, 2015); the function of ATM and ATR within the DNA damage response will be discussed here.

In humans, the Ataxia telangiectasia (AT) syndrome led to the discovery of the protein ATM (AT mutated), with AT patients showing symptoms of immunodeficiency, cancer predisposition, neurodegeneration and genome instability, which are strongly linked ATM to genomic maintenance. Cells that lack ATM fail to respond to DNA damage induced by IR, with documented absence of checkpoint activation (Bakkenist and Kastan, 2003). ATR was first identified via analysis of the human genome database by sequence similarity with ATM and *S. pombe* Rad3 (Cimprich et al., 1996). In contrast to ATM, gene disruption of ATR has been shown to be embryonically lethal in mice with partial loss of function causing severe phenotypes, indicating non-redundant activities of ATM and ATR (Brown and Baltimore, 2000; de Klein et al., 2000; O'Driscoll and Jeggo, 2006).

ATM is well known as the pivotal activator of the DNA damage response on detection of a DSB, orchestrating a global response including checkpoint activation, DNA repair, senescence, apoptosis and changes to chromatin structure and transcription (Shiloh and Ziv, 2013). With such a wide range of responses, the activity of ATM must be tightly regulated to abate spurious activation. In un-perturbed cells ATM remains in an inactive state, trans-dimerised via its kinase domain. Upon DNA damage caused by IR, ATM is at the forefront of the detection with rapid recruitment and activation upon DSB formation (Giunta et al., 2010; Lee and Paull, 2005). Although originally this highlighted ATM as a sensor of DNA damage, more recent evidence shows that the MRN complex (Mre11, Rad50, and Nbs1) and Ku (Ku70-Ku80) are the sensors of the DSB subsequently recruiting ATM for activation, providing another level of ATM regulation (Blackford and Jackson, 2017).

Although this area is still widely debated, a complex interplay between sensors and activation has become apparent. Mice knockout studies using the Cre/LoxP recombination system have shown compensation between the MRN and Ku70/Ku80 recruitment pathways with activation of ATM in the absence of MRN or Ku70/Ku80 (Hartlerode et al., 2016). ATM has also been shown to have MRN/DSB independent methods of activation in response to oxidative stress or chromosome changes (Blackford and Jackson, 2017).

The exact method in which recruitment activates ATM is however yet to be fully understood. It has been shown that intramolecular phosphorylation of ATM disrupts the dimer allowing activation (Bakkenist and Kastan, 2003). *In vitro* kinase assays have also shown activation of ATM by MRN using purified components, stimulating the kinase activity towards its substrates (Lee and Paull, 2005), with the potential for a positive feedback loop to reinforce ATM activation (Kaidi and Jackson, 2013).

After recruitment and activation, ATM acts upon its numerous targets including BRCA1 (Cortez et al., 1999), CtIP (Li et al., 2000), and itself (Bakkenist and Kastan, 2003) to bring about cellular changes. The first target at a DSB damage site is the C-terminal tail of the histone variant H2AX; the phosphorylated

H2AX binds MDC1 (mediator of DNA damage checkpoint protein-1) leading to further MRN recruitment (Lloyd et al., 2009). A cascade of recruitment follows with MRN bringing further ATM to amplify the formation of γ-H2AX (Harper and Elledge, 2007). This amplifies the signal ensuring recruitment and maintenance of proteins such as MRN, BRCA1 and 53BP1 at the damage site (Stucki and Jackson, 2006). This activation is summarised in figure 1.3.3. More recently, the activation of ATM by MRN has been shown to phosphorylate the MRN complex to control the extent of break resection and formation of ssDNA (Kijas et al., 2015; Lavin et al., 2015). Notably ATM also phosphorylates kinases such as Chk2 (Matsuoka et al., 1998) and MAPK (Mitogen activated protein kinase) (Vanesa Lafarga et al., 2009) to act on yet more substrates (Shiloh and Ziv, 2013).

One such substrate is the tumour suppressor protein p53. p53 Can be indirectly activated through the MAPK pathway as well as through ATM phosphorylation of MDM2 (Mouse Double Mutant -2) and p53 co-factors (Blackford and Jackson, 2017; Shiloh and Ziv, 2013). At the same time, ATM can phosphorylate p53 on multiple sites to stabilise the protein by resistance to MDM2 degradation (Ashcroft et al., 1999; Banin et al., 1998; Canman et al., 1998). Together, all these actions of ATM orchestrate global cell changes to activate cell-cycle arrest and allow time for DNA repair.

In contrast to ATM, ATR acts as the pivotal kinase during replication stress and is activated by a much larger variety of genotoxic stress, primarily UV radiation and replication inhibitors (Ward et al., 2004). ATR has been shown to be recruited to extended tracks of ssDNA, which occur from nucleolytic processing of various forms of DNA damage including DSBs or by uncoupling of the helicase-polymerase complex at replication forks (Byun et al., 2005). Both of these events cause rapid coating of the ssDNA by Replication protein A (RPA) which then recruits ATR via its ATR interacting protein (ATRIP) subunit (Zou and Elledge, 2003). Studies using *Xenopus* egg extracts have shown that the depletion of RPA completely inhibits the chromatin association of ATR; ATRIP was also shown necessary for controlled recruitment of ATR to RPA (Zou and Elledge, 2003). However, recruitment of ATR is insufficient for optimal activation, with further proteins being necessary, such as TopBP1 (topoisomerase binding protein 1) which is a crucial activator of ATR (Kumagai et al., 2006). As with ATR, loss of TopBP1 or inability to activate ATR has been shown to be lethal in mammalian cells (Jeon et al., 2011; Zhou et al., 2013). The interaction between ATR-ATRIP and TopBP1 is also necessary to supress new origin firing (Sokka et al., 2018). Potentially overlapping mechanisms exist for recruitment of TopBP1 to ATR that implicate factors such as MRN and Bloom syndrome protein (BLM) (Blackford et al., 2015; Duursma et al., 2013). Thus, a prominent pathway for recruitment of TopBP1 and activation of ATR consists of the RPA-ssDNA binding, interaction with the RAD17-RFC (replication factor C) clamp loader complex (Zou and Elledge, 2003) and the 9-1-1 (RAD9-HUS1-RAD1)

heterotrimer (Delacroix et al., 2007); TopBP1 is then recruited through sequential binding and interactions with the C-terminal tail of RAD9, leading eventually to activation of ATR (figure 1.3.3).

Like ATM, ATR has many phosphorylation targets. Some of the targets overlap, such as MRN which acts both up- and down-stream of ATR activation. For example, DNA-end resection by MRN causes recruitment of RPA leading to ATR activation; however, replication fork stalling following UV treatment causes ATR-mediated MRN phosphorylation to down regulate DNA replication (Olson et al., 2007). On activation, a key ATR function is to phosphorylate and activate the protein kinase Chk1, as shown to occur in *Xenopus* egg extract studies through the mediator protein Claspin (Kumagai et al., 2004). As demonstrated by mutational analysis, Chk1 is an essential kinase required for the G2/M checkpoint passage (Kumagai et al., 2004). On activation, Chk1 promotes proteasomal degradation of CDC25, a protein responsible for removing inhibitory phosphorylation events on CDKs. Thus, reducing CDK activity slows down or arrests cell-cycle progression, allowing time for DNA repair (Bartek et al., 2004), as well as inhibiting further late-origin replication firing and increasing the pool of nucleotides available for replication (Blackford and Jackson, 2017). Although many of the signalling events in the DDR rely on phosphorylation, other post-translational modifications also play a key role such as sumoylation, ubiquitination, acetylation, and methylation (Harper and Elledge, 2007; Polo and Jackson, 2011).

With ATM and ATR's pivotal roles in the DNA damage response, it is not surprising that a lot of research is underway focusing on ATM and ATR as possible targets for cancer treatment. Targeting and inhibiting the transduction signalling would inhibit propagation of the repair signal, therefore sensitizing cells to DNA damaging agents such as IR in combination therapy. Pre-clinical experiments suggest that inhibition of both proteins can selectivity sensitise cancer cells to radiation or chemotherapy (Dohmen et al., 2017; Fokas et al., 2014).

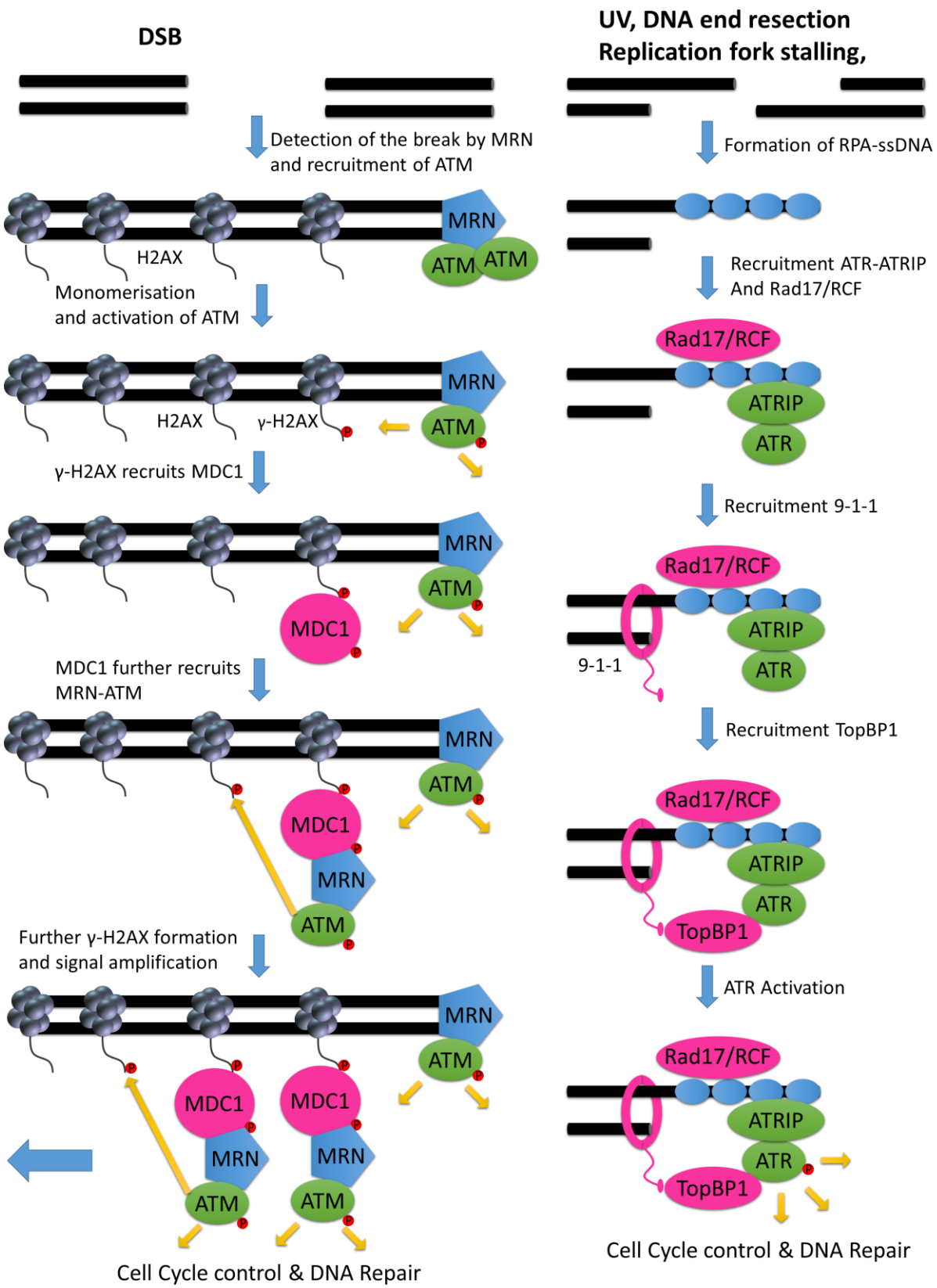


Figure 1.3.3: **Schematic representation of ATM and ATR activation.** Blue represents the sensor, green represents either ATM or ATR-ATRIP, with pink representing the mediator of activation. Yellow arrows indicate phosphorylation events with phosphates in red.

1.3.4 The ATM/ATR checkpoint signalling cascade

Interplay between the DDR and cell cycle machinery results in distinctive signalling in response to DSBs in different stages of the cell cycle for checkpoint activation and recovery back into the cell cycle. Once activated, ATM and ATR phosphorylate a multitude of proteins including those with roles as checkpoint mediators such as BRCA1, 53BP1, MDC1 and claspin. Direct or indirect phosphorylation of the checkpoint transducers Chk1 and Chk2 by ATM and ATR respectively is followed by a further cascade of events, with effectors such as CDC25 and p53 leading to cell cycle arrest (figure 1.3.4) (Blackford and Jackson, 2017; Sancar et al., 2004). Details surrounding the G1-S, intra S and G2-M phase checkpoint activation by ATM/ATR will be briefly described in the following section.

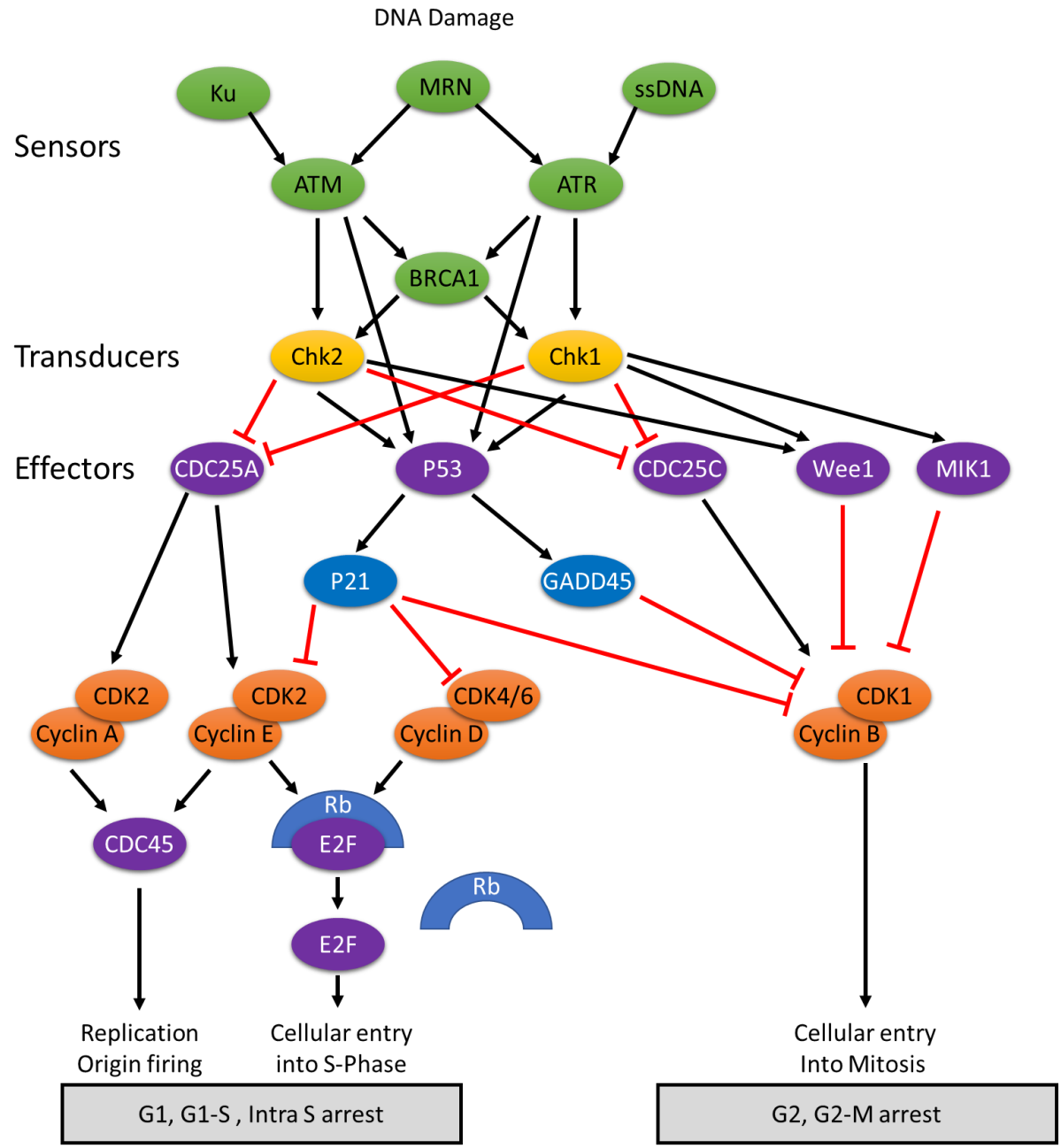


Figure 1.3.4: **ATM and ATR activated checkpoint pathways.**

1.3.4.1 G1-S phase checkpoint

In the context of DSBs, ATM is the activator of the G1-S checkpoint. During the G1 phase the main repair pathway is Non-Homologous End Joining (NHEJ) which has insufficient DNA-end resection to activate ATR (Cortez et al., 2001; Jazayeri et al., 2006). Conversely, small-molecule inhibition of ATR has been shown to allow early entry into S-phase with DNA damage present, implying that ATR still has an important function in G1-S checkpoint (Bøe et al., 2018). After detection of DNA damage, Chk1 and Chk2 phosphorylate the CDC25A phosphatase, which through binding to 14-3-3 is sequestered to the cytoplasm and unable to activate Cyclin E- CDK2 and therefore maintains cells in G1-phase (Chen et al., 2003).

Activation of ATM/ATR, Chk1/Chk2 leads to phosphorylation and stabilisation of p53, which in turn lead to transcription of a wide variety of targets including p21, a CDK inhibitor protein (Harper et al., 1995; Wade Harper et al., 1993). P21 inhibits the S-phase promoting Cyclin E-CDK2 thus maintaining G1 arrest, along with inhibition of cyclin D-CDK4/6 preventing phosphorylation of Retinoblastoma (RB). Both targets are crucial to progression into S-phase. Un-phosphorylated RB remains bound to the E2F transcription factor preventing transcription of S-phase genes for the G1-S phase transition (Sancar et al., 2004). This process has been reported to oscillate in waves of ATM, p53 and p21 activation until the DSB is resolved, allowing the cell a simple pathway for progression and removal of the cell cycle arrest on DNA repair (Batchelor et al., 2009). Conversely, when the DNA damage is too severe ATM activation of the MAP kinase family leads to a slower transcriptional response; this allows stabilisation of p21's mRNA to sustain the cell cycle arrest (V. Lafarga et al., 2009; Reinhardt et al., 2007).

1.3.4.2 S-phase checkpoint

While ATM checkpoint activation occurs in S-phase the main checkpoint activator is ATR. Once replication has been initiated by CDK2, a DSB can trigger activation of ATR through end resection of the DNA during Homologous Recombination (HR) repair. As mentioned earlier, resection of DNA exposes ssDNA, which through RPA activates ATR, mediating enhanced downstream checkpoint signalling to allow time for HR repair. Genotoxic stress can also arise from other forms of DNA damage causing replication forks to stall, thus also activating ATR, with the need to inhibit further origin firing. Alongside the actions of ATR to further stabilise p53 and the p21 pathway a further protein is involved, the Wee1 kinase. Wee1 is phosphorylated by Chk1 and Chk2; phosphorylation activates the kinase to phosphorylate CDK2 and inhibit its activity (Beck et al., 2010). During S-phase p21 acts at DSBs. However, due to its role in origin licencing it is degraded at the replication fork by the proliferating cell nuclear antigen (PCNA) associated $CRL4^{cdt2}$ ubiquitin ligase to inhibit detrimental re-replication (Y. Kim

et al., 2008). ATR also targets the Fanconi Anaemia (FA) pathway, responsible for inter-strand crosslink repair, by phosphorylating proteins FANCI and promoting FANCD2 mono- ubiquitination, which are key steps in the pathway activation (Ishiai et al., 2008).

1.3.4.3 G2-M phase checkpoint

The G2-M checkpoint is crucial for preventing cells from entering mitosis in the presence of damaged DNA. In M-phase the DDR is potently inhibited (Shaltiel et al., 2015). DNA damage sustained in this phase activates ATM but the signal is not transmitted to activate Chk2 (Giunta et al., 2010). Therefore, cells must ensure entrance into M-phase is delayed in the presence of DNA damage. The Wee1-dependant pathway of CDK inactivation remains important in G2 phase alongside a second kinase MIK1 which acts on the same site in CDK1 enforcing the inactive state (Baber-Furnari et al., 2000; Hirai et al., 2009). Similar to the previous two checkpoints, via ATR and ATM, Chk1 and Chk2 phosphorylate CDC25C leading to its exclusion from the nucleus and inability to dephosphorylate Cyclin B-CDK1 (Graves et al., 2001; Lopez-Girona et al., 1999). The actions of p21's present in G1 can now also be reinstated as replication has completed, and transcriptional control of p21 through ATM and the MAP kinase pathway can again cause sustained arrest (Bunz et al., 1998). In G2 phase, p53 can also play a further role through transcriptional expression of GADD45 (growth arrest and DNA damage-inducible gene 45α), which disassociates the cyclin B-CDK1 complex. In G1 the action of GADD45 does not cause arrest due to inefficient inhibition of the cyclin E-CDK2 complex (Jin et al., 2000; Zhan et al., 1999). In further contrast to G1, DSBs can now be faithfully repaired by HR due to the presence of the sister chromatid. Therefore ATR is activated in G2 and can maintain cell cycle arrest (Brown and Baltimore, 2003). Inhibition of HR repair proteins BRCA1 and PALB2 cause premature M-phase entrance, highlighting the importance of HR and ATR activation for G2 arrest (Menzel et al., 2011). Knockdown of mediator proteins such as 53BP1 and MDC1 also show a similar response as ATM activation at DSBs is not sustained (Shibata et al., 2010). Recovery from cell cycle arrest is a fine balance to avoid DNA damage persistence while avoiding senescence or apoptosis. Late activation of G2 arrest can also cause daughter cells to exit from the cell cycle or a delay in G1 (reviews in Shaltiel et al., 2015).

1.3.4.4 BRCA1 and CtIP in checkpoint pathways

Alongside proteins such as Chk1, Chk2 and p53 many other proteins are activated in the G1-S, S and G2-M DNA damage checkpoints. BRCA1 is an example of a protein phosphorylated at several sites by ATM, Chk2 and ATR, which leads to a cellular checkpoint response (Roy et al., 2011). In particular Chk2 phosphorylation of BRCA1 is required for the CtIP/MRN/BRCA1 complex formation to activate the G2/M checkpoint (Greenberg et al., 2006). Downstream of H2AX phosphorylation by ATM, BRCA1 has also been reported to be retained at chromatin resulting in the activation of Chk1 and the G2-M checkpoint (Huen et al., 2010). Studies have also shown the BRCA1/CtIP complex is necessary for ATR activation downstream of ATM, likely due to the role of CtIP in DNA-end resection (Sartori et al., 2007; Yu et al., 2006). BACH1 binds to the same site on BRCA1 as CtIP with the two interactions being mutually exclusive, forming the BRCA1 BARD1-BACH1-TopBP1 complex. This second complex has been shown to play a checkpoint regulatory role critical for the S-phase checkpoint activation (Greenberg et al., 2006). On IR or UV treatment the CtIP-BRCA1 interaction is necessary for the ATR/ATM dependant phosphorylation of Chk2, p53 and Nbs1, and downstream activation of P21 and GADD45 (Foray et al., 2003; Li et al., 2000, 1999). BRCA1 has also been shown to induce apoptosis by playing a role in regulation of anti-apoptotic and pro-apoptotic factors in response to DNA damage: upregulation of BRCA1 facilitates stress-induced apoptosis in cancer cell lines and increases susceptibility to DNA damage induced apoptosis (Thangaraju et al., 2000; Ting and Lee, 2004).

1.3.5 DNA damage induced apoptosis and its therapeutic potential

DNA repair defects, alongside irreparable, or intolerable DNA lesions threaten the genome stability of an organism. Programmed cell death, termed apoptosis, can be induced by DNA damage or natural processes to mitigate the effects of these events throughout the cell cycle. The main process of apoptosis involves activation of caspase (cysteine aspartyl protease) 3 leading to irreversible DNA fragmentation and apoptotic body formation (Elmore, 2007; Norbury and Hickson, 2001). Caspase 3 can be activated by external signalling through DEATH receptors or through an internal shift between pro- vs anti-apoptotic factors leading to the release of cytochrome C from the mitochondria and caspase 3 activation (Elmore, 2007; Fulda and Debatin, 2006; Wajant, 2002; Wang and El-Deiry, 2003).

The DDR is directly linked to apoptosis through p53 (Reinhardt and Schumacher, 2012). Sustained p53 activation directly activates transcription of pro-apoptotic genes while suppressing anti-apoptotic protein activity (De Zio et al., 2013; Elmore, 2007). Apoptosis can also occur through Chk2 via two pathways: E2F-1 transcriptional regulation of apoptotic factors and phosphorylation of PML (promyelocytic leukaemia) to promote pro-apoptotic activity (Stevens et al., 2003; Yang et al., 2002). Activation of p21 and p16 by p53 can also cause senescence. Senescence is a more permanent form of cell cycle arrest which also protects the organism during genotoxic stress (Althubiti et al., 2014).

The balance of apoptosis is a key factor in many human conditions including autoimmune disorders, neurodegenerative diseases and many types of cancer. However, the ability to modulate apoptosis has great therapeutic potential. Inhibition of the cell cycle checkpoints can lead to cells progressing to mitosis with damaged DNA causing mitotic catastrophe and cell death, selectively targeting replicating cells (Bouwman and Jonkers, 2012; Curtin, 2012). Inhibition of the cyclin-CDK complex is therefore a target for cancer treatment. In 1992, Flavopiridol was the first wide range CDK inhibitor shown to act as an anti-cancer agent, however with toxic side effects (Sausville, 2002; Visconti et al., 2016). More recently, selectivity for CDK4/6 to cause G1-phase arrest has been achieved with palbociclib, approved for the treatment of advanced breast cancer (Finn et al., 2015; Pernas et al., 2018).

Chemotherapeutic modulation of DNA repair pathways can also lead to apoptosis. Synthetic lethality can be employed when cancer cells are deficient in a DNA repair pathway. This is apparent by the sensitivity of BRCA-deficient tumour cell lines to PARP (poly ADP-ribose polymerase) inhibition, specifically targeting tumour cells to apoptosis, while the cells with intact BRCA1/2 pathways can repair the DNA damage (Audeh et al., 2010). Knowledge of how the different pathways act, interact and overlap can lead to further advances within the chemotherapeutic field and our understanding of human disorders. Research into DNA repair is therefore vital with the current understandings of the pathways discussed within the next sections of this thesis.

1.4 DNA Repair pathways

The first experimental evidence of what was DNA repair emerged in 1949 from observations made by Albert Kelner, who at the time was using UV radiation to generate mutations in bacteria. Kelner noted highly increased rates of survival when the irradiated bacteria were exposed to visible light (Kelner, 1949). At the same time, Renato Dulbecco discovered the phenomenon of photoreactivation in bacteriophage (Dulbecco, 1949). The correct interpretation of Kelner and Dulbecco's findings as a biochemical pathway for the repair of DNA photoproducts had to wait for the recognition of the DNA double helix as the chemical basis of genetic heredity in 1953, and the subsequent isolation of photolyase, the enzyme that catalyses photoreactivation in 1958 (Rupert et al., 1958). In 1974, Francis Crick wrote: "We totally missed the possible role of enzymes in repair, although, due to Claud Rupert's early very elegant work on photoreactivation, I later came to realize that DNA is so precious that probably many distinct repair mechanisms would exist." (Crick, 1974). By the 1960s, DNA repair had become an important and established field of research in molecular and cellular biology (Friedberg, 2008).

There are many different pathways of DNA repair, which are specific for the different types of DNA damage. In 2015, Tomas Lindahl, Aziz Sancar and Paul Modrich shared the Nobel Prize for Chemistry for their discoveries of Base Excision Repair (BER), Nucleotide Excision Repair (NER), and Mismatch Repair (MMR) respectively, which form the main pathways for repair of single-strand DNA lesions. Whereas DNA repair by direct reversal requires a single enzyme, NER, BER and MMR have multiple steps involving excision of the damaged region, insertion of new nucleotide(s) and ligation to complete the repair process (Clancy, 2008; Friedberg, 2008). Although the underlying mechanistic principles are similar and involve a sequential cascade of enzymatic steps, separate sets of proteins are at play in each repair pathway.

1.4.1 Direct Reversal

The first historical evidence of DNA repair can be explained by the repair mechanism of enzymatic photo-reactivation (EPR) which is also believed to be the first DNA repair mechanism to have evolved (Friedberg et al., 2005). Cyclobutane pyrimidine dimers are known to be one of the most common effects of UV radiation and have the potential to block both replication and transcription of DNA. They can however be repaired by DNA photolyases which utilise the energy of blue light to split the cyclobutane ring, thus directly reversing the DNA damage.

Another example of direct reversal is repair of alkylation affecting either the aromatic base or DNA backbone. The first example of base alkylation is N^3-methyladenine which can block replication and is removed directly by 3-methyladenine-DNA-glycosylases restoring the natural adenine base (Borja-Cacho and Matthews, 2008). Another example is that of O^6-methylguanine which incorrectly pairs during replication with thymine rather than cytosine and is therefore mutagenic. This lesion can be exploited in chemotherapy using the drug Temozlomide to generate O^6-methylguanine in replicating cells; only a single round of treatment can be performed due to drug resistance from enhanced activity of O^6-methylguanine-DNA-methyltransferase (MGMT) which directly reverses these lesions (Abbott and Saffhill, 1979; Fan et al., 2013).

1.4.2 Base Excision Repair

Base modifications are one of the most common forms of DNA lesion, accounting for thousands of DNA damage events per human cell per day (Hoeijmakers, 2009). Base Excision Repair (BER) has evolved to cope with the high level of spontaneous DNA damage from oxidation, deamination and alkylation. Although there is little distortion to the structure of the DNA helix, the damage can be recognised by DNA glycosylases that cleave the N-glycosidic bond between the damaged base and the ribose of the sugar-phosphate backbone, leaving an apurinic/apyrimidinic (AP) site for further processing. At least 11 distinct mammalian DNA glycosylases are known, each recognising a subset of base lesions with partial overlap between the enzymes. After generation of an AP site the repair can be completed by one of two pathways: Short-Patch (also known as single-nucleotide) BER or Long-Patch BER; the latter replaces between 2 to 13 nucleotides (Fromme and Verdine, 2004; Kim and Wilson III, 2012).

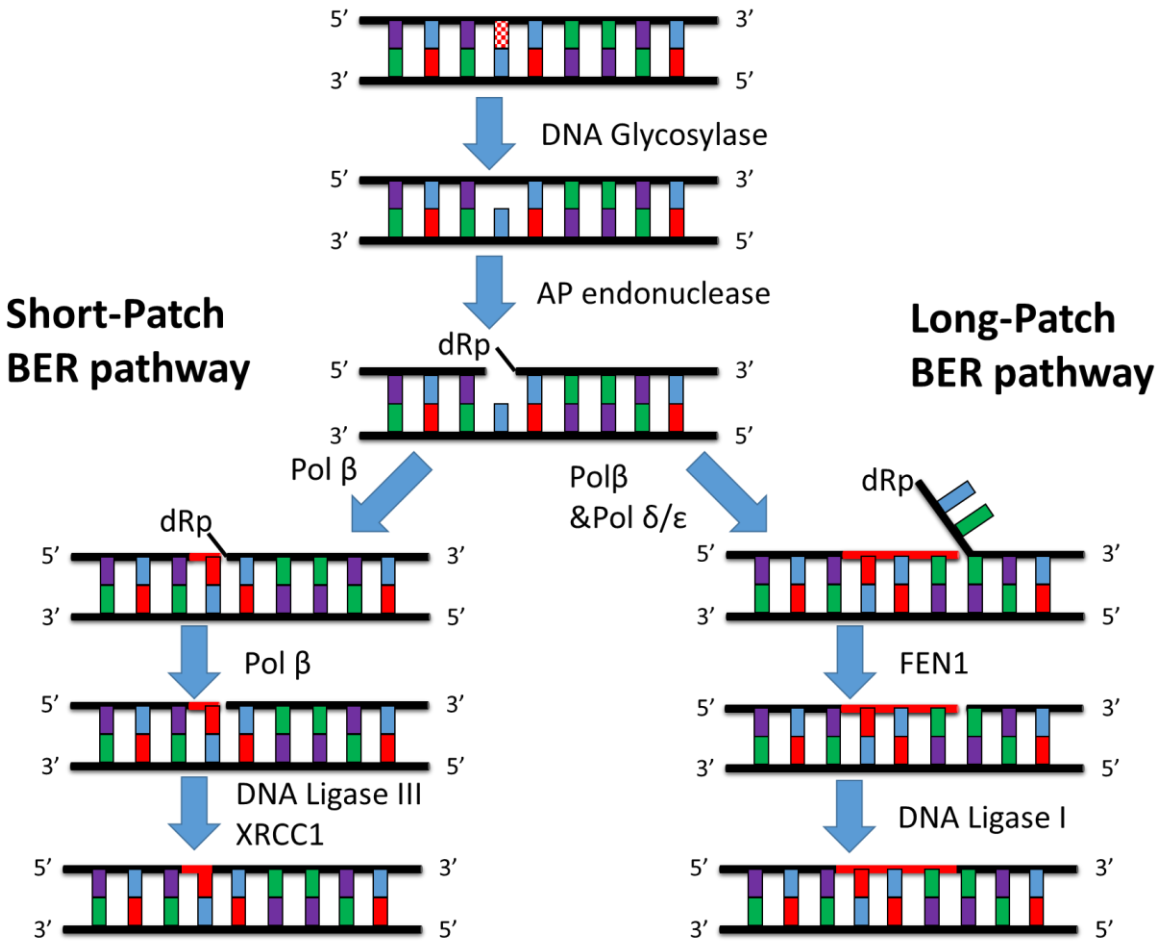


Figure 1.4.2**: A schematic drawing of the Short-patch and Long-Patch Base Excision Repair (BER) pathways for repair of DNA base lesions.**

Short patch BER is generally the dominant pathway in replicating and non-replicating cells. Long patch BER mainly takes place post-replication in proliferating cells utilising the proteins of DNA replication for processing after the generation of the AP site. The main step in short patch BER is cleavage of the phosphodiester bond 5' to the AP site by AP endonuclease APE1, which generates a strand break with a 3'-hydroxyl and 5' sugar phosphate end (5'dRp). DNA polymerase (Pol) β can then add one correct nucleotide to the nick while its phosphodiesterase activity (dRpase) removes the 5'dRP to generate a 5' phosphate end required for ligation. DNA ligase III-XRCC1 seals the break and completes the process (Krokan and Bjoras, 2013; Svilar et al., 2011). In Long patch BER the same initial process takes place, however Pol δ/ε replaces Pol β to displace several nucleotides extending the repair patch. This generates a flap with a 5'-sugar phosphate which can be removed by the FEN1 endonuclease. Following removal of the flap, DNA ligase I seals the break to complete the process (figure 1.4.2). Other proteins also have roles in BER to signal the DNA damage and to edit the ends of the nucleotide gap (Krokan and Bjoras, 2013; Memisoglu and Samson, 2000). PARP also plays a role in initial recruitment of BER proteins to damaged DNA by creating a scaffold for protein binding; inhibition of this pathway results in processing by the slower homologous recombination mediated pathways (Javle and Curtin, 2011).

1.4.3 Nucleotide Excision Repair

Nucleotide Excision Repair (NER) detects and removes bulky helix-distorting adducts by targeting altered DNA backbone conformations (Kim and Wilson III, 2012). The process was originally discovered through the process of bacteria repair of cyclobutene pyrimidine dimers (Hanawalt, 1994; Setlow, 1966), with a preference to repair transcribed genes compared to non-transcribed genes (Mellon et al., 1986). Our understanding of NER has been ever advancing through study of such conditions as Cockayne syndrome and Xerodema Pigmentosum (Hoeijmakers, 2009). From probing these conditions, looking at both transcribed and non-transcribed regions of DNA, two sub-pathways of NER were determined: Global Genome NER (GG-NER) and Transcription Coupled NER (TC-NER). Defects in these pathway lead to a variety of diseases from UV radiation sensitivity to severe premature aging (Marteijn et al., 2014).

The sensing of DNA lesions for the two sub-pathways involves a different subset of proteins. For GG-NER the damage sensor XPC (in complex with HHR23B) constantly probes the DNA for helix-distorting lesions (Sugasawa et al., 1996). In TC-NER damage is directly recognised by the stalling of RNA pol II during transcription elongation. Transient interactions with NER proteins allow quick recognition and increased binding affinity of proteins such as Cockayne Syndrome Protein (CSB) to RNA pol II (Marteijn

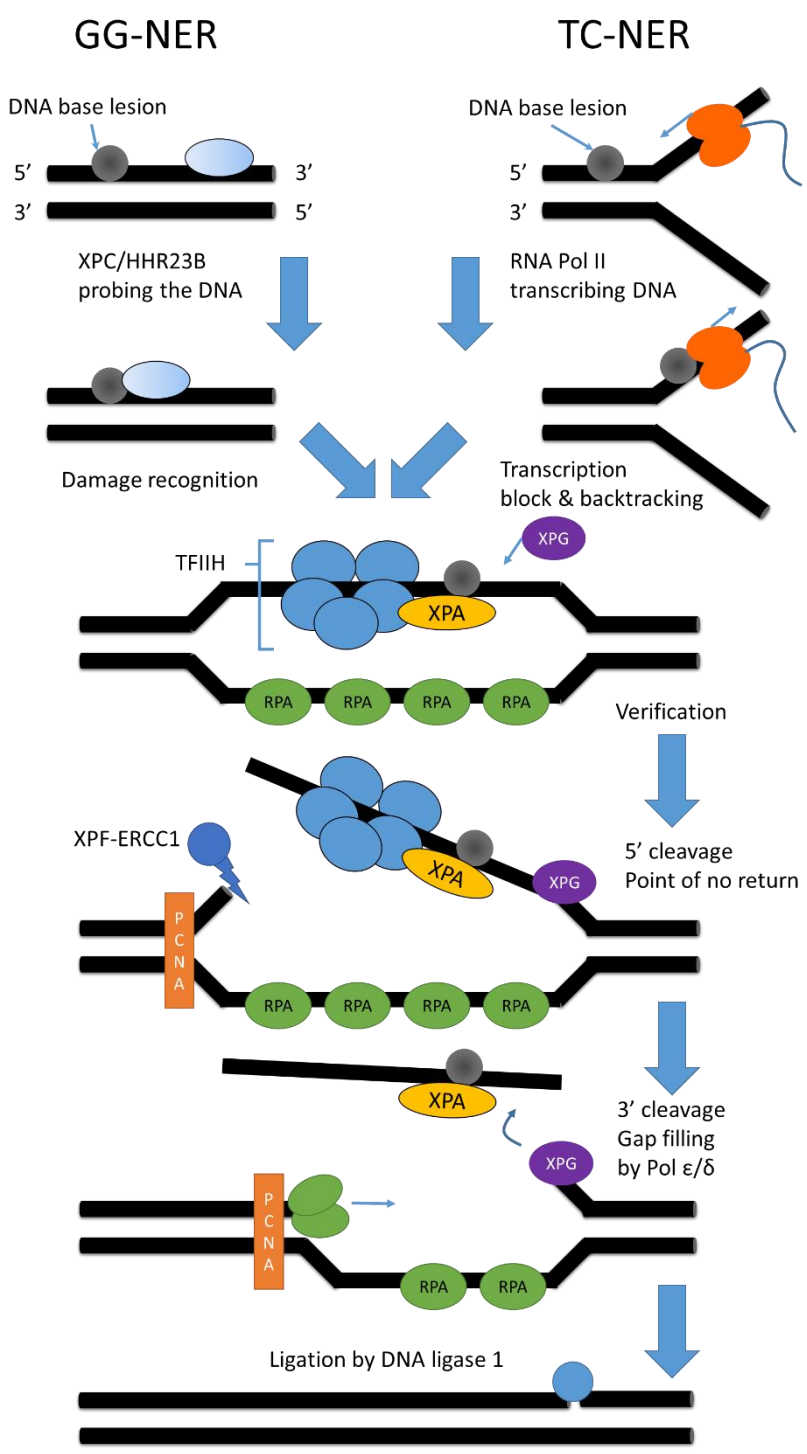


Figure 1.4.3: **A schematic diagram of DNA base lesion repair by Global Genome Nucleotide Excision Repair (GG-NER) and Transcription Coupled Nucleotide Excision Repair (TC-NER).**

et al., 2014). The subsequent recruitment of RPA, XPA and the 11 subunit Transcription initiation factor IIH (TFIIH) complex forms the pre-incision complex 1 (PIC1) which is the same in both pathways. The TFIIH helicase activity further unwinds the DNA allowing for recruitment of XPG which displaces XPC to form pre-incision complex 2 (PIC2). XPG can then verify the existence of lesions by binding to chemically altered nucleotides. Once verified, pre-incision complex 3 (PIC3) is formed by recruitment of the XPF-ERCC1 nuclease which is directed to the damaged stand by RPA to create an incision 5' to the lesion. After this "point of no return", XPG is activated cleaving 3' to the lesion and creating a 22-30 nucleotide long strand. The gap can then be filled by DNA Pol ε/δ and PCNA before ligation by DNA ligase 1 (Adebanke F. Fagbemi. et al., 2011; de Boer and Hoeijmakers, 2000; Orelli et al., 2010; Sugasawa et al., 2009). Some NER factors are known to be involved in other processes such as TFIIH in transcription, XPF-ECC1 in recombination and RPA in replication and recombination. The two pathways are summarised in figure 1.4.3.

1.4.4 Replication fidelity and Mismatch Repair (MMR)

Preservation of genetic information during normal cell proliferation relies on low mutation rates in DNA replication. High fidelity DNA polymerases maintain genomic integrity by ensuring that only the correct Watson-Crick base pairing occurs between the templated base and the nucleotide in its active site. Studies on Pol δ show only one incorrect base pair per 10^7 to 10^8 nucleotides inserted (Bębenek and Ziuzia-Graczyk, 2018; Ganai and Johansson, 2016); however given the size ($3.2x10^9$ nucleotides) of the human genome multiple errors still occur. Part of this high fidelity is achieved through proofreading by Pol δ and Pol ε, which remove incorrectly inserted bases using their 3' to 5' exonuclease activity (Schmitt et al., 2009). Errors which escape this process are repaired by the highly conserved MMR, while mutations in the MMR pathway can lead to Hereditary Non-Polyposis Colorectal cancer (HNPCC) (Steinke et al., 2013).

In MMR, the cell must distinguish between the two strands to ensure removal of the mismatch from only the daughter strand. In bacteria strand discrimination exploits the transient hemi-methylation which lags behind the replication fork. For humans however the mechanism has been less clear and is likely to rely on nicks within the newly synthesised DNA (Hsieh and Zhang, 2017). Nicks are present on the lagging strand prior to ligation of okazaki fragments however this implies nicks are present on leading and lagging strands to direct the MMR machinery to the daughter strand. Orientation specific loading of PNCA has been reported on the leading strand providing evidence to support this method of detection (Pluciennik et al., 2010). Following mismatch detection, the excision machinery is recruited including EXO1 and RPA triggering strand degradation. RPA coats single stranded DNA prior to Pol δ recruitment for re-synthesis and the remaining nick sealed by DNA ligase I (Li, 2007; Thornby Bak et al., 2014).

1.4.5 Repair of interstrand crosslinks

DNA interstrand crosslinks (ICL) have the potential to cause replication and transcription stalling through blockage of DNA separation. Some chemotherapeutic drugs introduce ICL to promote cell death; agents such as cisplatin that introduce ICLs were one of the earliest and still widely used chemotherapeutics. However, cisplatin causes only 5% interstrand crosslinks, and over 90% intrastrand crosslinks, which can be repaired but are a source of further DNA damage leading to cancer (Deans and West, 2011; Muniandy et al., 2011). Fanconi Anaemia (FA) is a rare genetic disorder with high sensitivity to ICLs and its study has provided insight into the repair of these lesions, which requires a coordinated response involving proteins from NER, HR and the FA complex.

Initial ICL recognition is by the FANCM-FAAP23 complex which becomes part of the core FA complex. The formation of the FA complex can then coordinate changes in the fork such as regression and single

DSB formation to stabilise an intermediate structure. Recruitment of the XPF-ERCC1 complex can then create a gapped structure for lesion bypass polymerases to add nucleotides across the ICL. The excised DNA, still bound by the crosslink, can then be further processed and removed by NER. The whole fork can then be repaired by HR allowing replication to proceed; FANCD2 has been shown to interact with CtIP to help channel the DSBs intermediate into the HR pathway (Unno et al., 2014).

1.4.6 Tolerance to DNA damage

Even with the many DNA repair pathways at the cell's disposal, DNA damage is a random process and can interfere with ongoing cellular processes before being repaired (Friedberg, 2008). In the case of DNA replication, a blocked replication fork can lead to cell death. Thus, cells have evolved tolerance pathways such as Translesion DNA Synthesis (TLS) to bypass the damaged site. During TLS, when the replication machinery stalls at a DNA lesion the replicative DNA polymerase is exchanged for a specialised Y-family TLS polymerase, such as Pol eta, kappa and zeta (Goodman and Woodgate, 2013). These polymerases have an active site capable of accommodating distorted or damaged templates, allowing additional nucleotides to be added to the newly synthesised strand opposing the lesion (Sale et al., 2012). This lesion bypass by TLS allows fork progression and abates cell death, but at the expense of an increased mutagenic load, as TLS polymerases are low-fidelity. Control mechanisms therefore exist to ensure tight regulation of TLS polymerase activity (Sale, 2013).

1.5 Double strand break repair

A Double Strand Break (DSB) is one of the most detrimental forms of DNA damage. Even a single DSB causes a major threat to the genome integrity by creating two uncapped chromosomal ends. Without repair chromosomal loss is possible, as the ends are accessible to nucleases, which can lead to cell death. Inaccurate repair can also be detrimental as large translocations or gene fusions can cause cell transformation and cancer (Jackson and Bartek, 2009; Salesse and Verfaillie, 2002).

DSBs can be caused by endogenous or exogenous sources (O'Driscoll and Jeggo, 2006). Endogenous DSBs can be a deliberate action in the execution of a cellular programme of genetic rearrangement. For example, during meiosis where DSBs are generated for the process of recombination between homologous chromosomes (Baudat and De Massy, 2007; Edlinger and Schlögelhofer, 2011). Development of immunity also relies on a controlled DSBs through V(D)J recombination and immunoglobulin heavy chain switching. In these processes segments of DNA in the immunoglobulin and T-cell receptor loci are shuffled and re-ligated to generate genetic diversity within the immune system (O'Driscoll and Jeggo, 2006; Schatz and Swanson, 2011).

Exogenous sources of DNA damage such as IR can also produce complex DSBs with multiple clustered sites of DNA damage. These can be produced either by generation of ROS that attack the phosphodiester backbone or by directly supplying the energy to break a bond. In many cases these events can be two SSBs on the complementary strands in close proximity to each other, which causes a DSB. Further DNA damage caused by IR such as protein adducts can increase the complexity of the break (Durante and Loeffler, 2010; Schipler and Iliakis, 2013; Ward et al., 1987). Failure of normal cellular processes, however, can also be a source of endogenous DSBs in an uncontrolled manor. As repair is not initiated in M-phase, prolonged mitosis from checkpoint failure can lead to an increased build-up of exogenous DNA damage. Physical pulling in opposite directions of the sister chromatids can also lead to further DSB formation (Ganem and Pellman, 2012). The replication fork can also lead to the formation of DSBs. On encountering a nick nascent strand synthesis can be discontinued leading to a DSB, replication fork stalling can also lead to the formation of DSB through collision of two forks (Aguilera and Gómez-González, 2008; Alexander and Orr-weaver, 2016). Furthermore, the non-translated ssDNA during transcription can form secondary structures that can impede fork progression and cause DNA breaks. Replication of DNA fragile sites can also cause secondary structures to be cleaved generating a DSB, with the replication fork meeting a SSB or fork stalling also leading to the formation of a DSB (Aguilera and Gómez-González, 2008).

Repair of a DSB is necessary to protect the cell against detrimental effects. Cells have therefore evolved two main mechanisms of DSB repair, NHEJ and HR.

1.5.1 NHEJ

Canonical NHEJ is a DSB repair mechanism without the need for resection and homology of the DNA. In NHEJ, the DSB ends are therefore directly re-joined with little or no base pairing at the damaged site. Although this can result in inaccurate repair, NHEJ is a major pathway of DSB repair in mammalian cells, whereas in lower eukaryotes such as *S. cerevisiae* recombination-mediated pathways predominate (Davis and Chen, 2013; Hefferin and Tomkinson, 2005).

The initial stage of NHEJ is recognition by binding of the Ku protein, a heterodimer of two highly abundant subunits within the cell Ku70 and Ku80. Ku quickly and tightly binds to DNA ends without sequence specificity forming a ring structure around the DNA. This has been further demonstrated by both crystal and electron microscopy structures of Ku in complex with dsDNA (Davis and Chen, 2013; Fell and Schild-Poulter, 2015; Spagnolo et al., 2006; Walker et al., 2001). These structures showed no obvious difference between a Ku DNA-bound and unbound state meaning after DSB repair it is likely that Ku remains trapped on the DNA (Rivera-Calzada et al., 2007). Removal of Ku was subsequently shown to be by protein ubiquitination and degradation by the E3 ubiquitin ligase RING finger protein 8 (RNF8) (Feng and Chen, 2012). Ku has also been reported to be involved in bridging the broken DNA ends. Knockdown of Ku80 showed an increased distance between the ends with Ku70 also shown to be involved in bridging interactions (Davis and Chen, 2013; Ribes-Zamora et al., 2007; Soutoglou et al., 2007).

Following binding of Ku further proteins are recruited such as the DNA-dependent protein kinase catalytic subunit (DNA-PKcs), X-ray cross complementing protein 4 (XRCC4), XRCC4-like factor (XLF) and DNA Ligase IV (Davis and Chen, 2013). Initial recruitment of DNA-PKcs ensures the Ku complex is at the tip of the DSB, with DNA-PKcs involved in bridging across the gap allowing auto phosphorylation and formation of the active holoenzyme with Ku. Further recruitment of XRCC4-DNA ligase IV complex is achieved through interactions with the Ku-DNA-PKcs complex with XRCC4 binding to Ku70 specifically (Chang et al., 2017; Davis and Chen, 2013). The XRCC4-DNA ligase IV complex can then complete the break ligation in an ATP-dependant manner; the protein XLF was identified to prompt this reaction (Ahnesorg et al., 2006; Dahm, 2008). More recently PAXX (PAralog of XRCC4 and XLF) has been identified as a further core NHEJ factor for mediation of XRCC4-DNA Ligase IV ligation to complete NHEJ (Ochi et al., 2015). This pathway is summarised in figure 1.5.1.

While the proteins described above have critical roles in the NHEJ pathway there is considerable flexibility in the proteins recruited for repair. Complex DSB ends can be produced by IR with a subset being processed by NHEJ. These however require further processing through kinases/phosphatases such as the mammalian polynucleotide kinase/phosphatase (PNKP) (Chappell et al., 2002), nucleases

such as Mre11 (Foster et al., 2011) and Artemis (Ma et al., 2002), Werner syndrome protein (WRN) a nuclease and helicase (Karmakar et al., 2002), the RECQ1 helicase (Parvathaneni et al., 2013), phosphodiesterases such as tyrosyl DNA phosphodiesterase 2 (TDP2) (Gómez-Herreros et al., 2013) and DNA polymerases such as Pol μ and Pol γ (Ma et al., 2004; Mahajan et al., 2002). For example, incompatible 5' overhangs require processing by Artemis in complex with DNA-PKcs for removal of the overhanging DNA, blunt end ligation can then be stimulated by PAXX and XLF leading to a deletion of information (Chang et al., 2017). All these factors process the DSB end to enable re-ligation of a variety of complex DSBs. This processing can result in a diverse range of products including insertions, deletions and translocations. However, this potentially error prone NHEJ repair is preferred to the potential loss of chromosome without DSB repair.

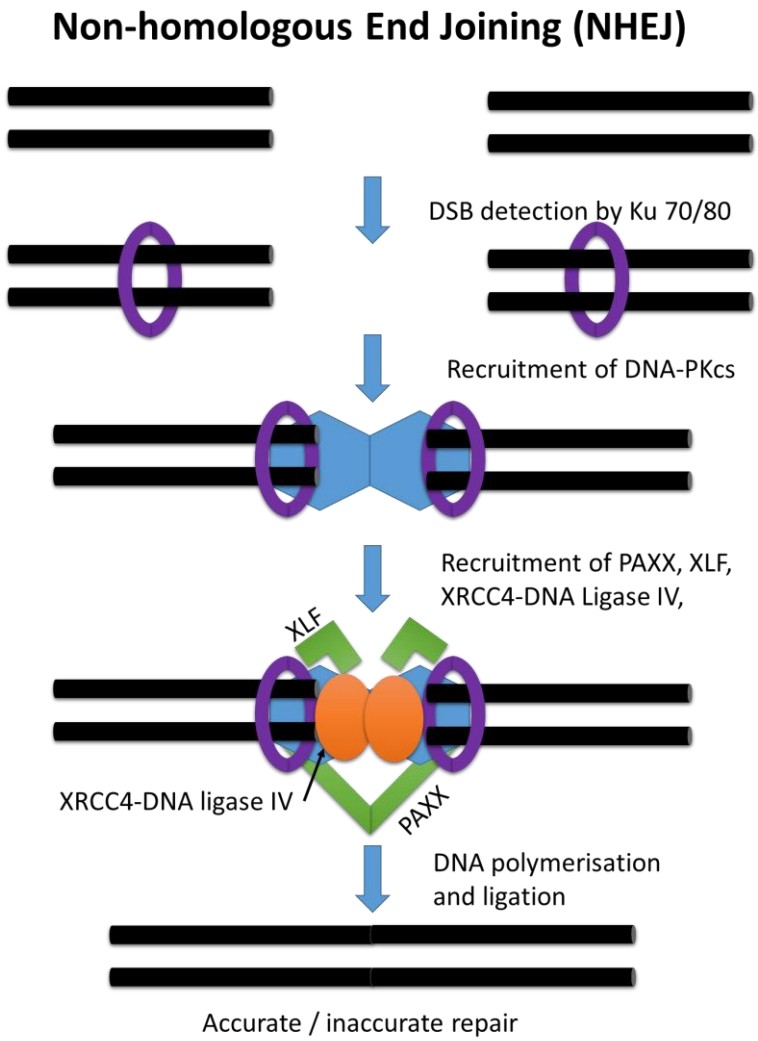


Figure 1.5.1: **Schematic representation of Non-Homologous End Joining.**

1.5.2 MMEJ

Micro homology mediated end joining (MMEJ) is an alternate, mutagenic, NHEJ DSB repair pathway and always results in deletions, potentially causing chromosome translocations to quickly repair DSBs. The MMEJ pathway is relied upon when NHEJ or HR are compromised while other mechanisms channel breaks away from MMEJ (Reczek et al., 2016). Due to the mutagenic property this pathway was thought to be a last resort to avoid apoptotic events (Sfeir and Symington, 2015). However, timely DSB repair is important for genomic integrity with evidence supporting a role for MMEJ within the cell in the presence of active HR and NHEJ pathways (Truong et al., 2013). CtIP has been established to play a role in MMEJ through DNA-end resection at a DSB. This resection inhibits the binding of Ku, therefore the break can no longer be processed by canonical NHEJ, in the absence of long range resection HR also is unable to process the break leading to MMEJ (Wang and Xu, 2017).

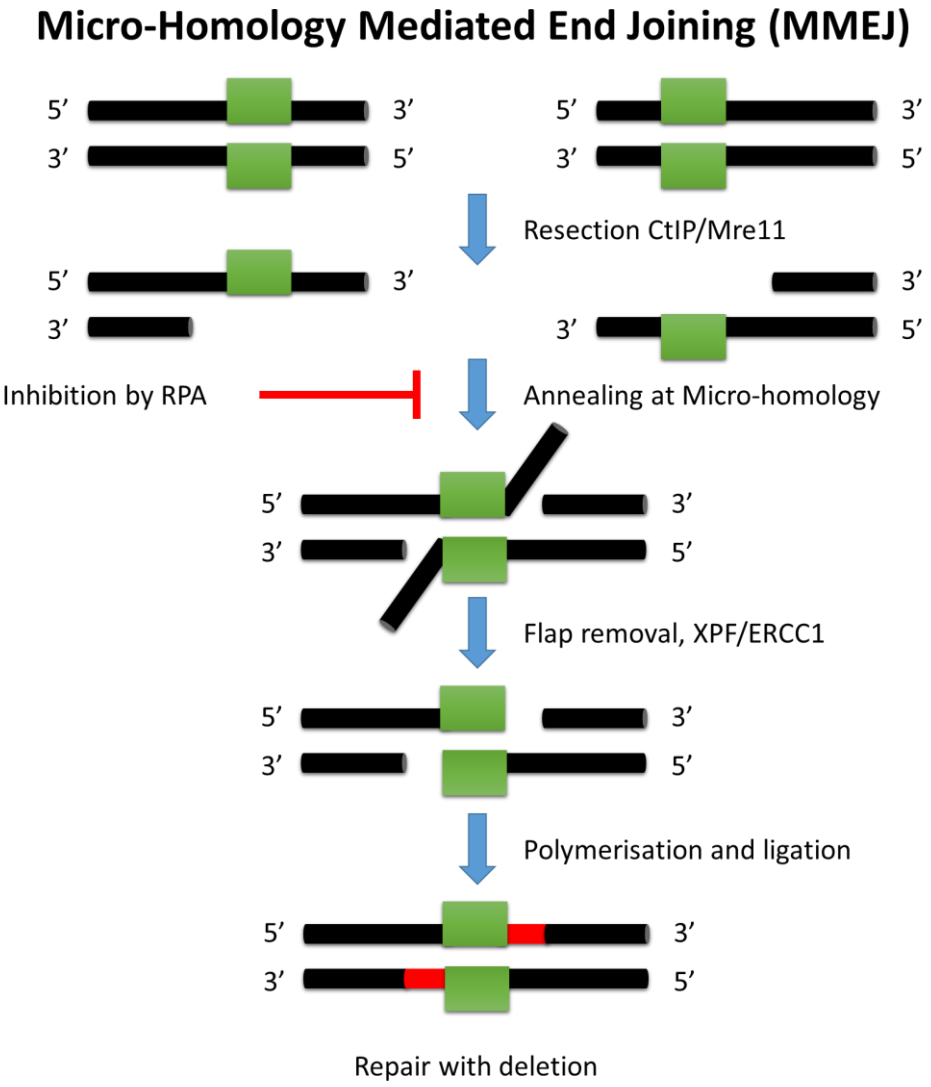


Figure 1.5.2: **Schematic representation of micro-homology mediated end joining.** Green represents a region of micro-homology necessary for MMEJ.

MMEJ is distinguished from NEHJ by the use of a 5-25 base pair micro-homology sequence between the broken ends independent of Ku. This resection stage is shared with HR, however it is BLM-, Exo1- and Rad51-independent, separating MHEJ from the HR pathway leading to a deletion (Truong et al., 2013; Wang and Xu, 2017). MMEJ is still an emerging pathway within DNA repair, however the key step of initial Mre11/CtIP dependent resection was solidified early in its discovery (Bennardo et al., 2008). Following end resection by CtIP and Mre11 the micro-homology region is annealed by an unknown mechanism, this process has been shown to be countered by binding of RPA (Wang and Xu, 2017). Annealing leads to formation of an intermediate with a 3' flap and gaps either side of the break site (figure 1.5.2). These 3' flaps must then be removed by XPF/ERCC1 to allow a promiscuous DNA polymerase pol θ to fill the gaps (Mateos-Gomez et al., 2015; Seol et al., 2018). PARP1 along with XRCC1 DNA ligase III/I have been implemented in this stage and the subsequent ligation however the precise mechanism is unclear (Audebert et al., 2006; Seol et al., 2018; Simsek et al., 2011).

Although MMEJ is mutagenic it has been seen that inactivation through inhibition of Pol θ results in vast chromosome deletions, showing its necessity for genome protection (Yousefzadeh et al., 2014). This role of Pol θ is crucial in DSB repair in a HR deficient background, with reports of Pol θ dependency in tumours that lack the HR repair pathway (Ceccaldi et al., 2015). Initiation of MMEJ is focused on CtIP converging cellular signals. This includes phosphorylation by polo-like kinase 1 (PLK1) expressed in S through to M phase. Action of PLK1 has been shown to inhibit CtIP/Mre11 resection sufficient for HR and channel the break into MMEJ; this action accelerates DNA repair through the quicker MMEJ and allows the G2/M transition in cells recovering from DNA damage (H. Wang et al., 2018). With CtIP playing a pivotal in the MMEJ pathway, inhibition of this action could be route to a further cancer treatment allowing cells to enter M-phase with DNA damage and cause mitotic catastrophe when other pathways are compromised. Nevertheless, further study into both CtIP and MMEJ is necessary to fully understand the process and the specific domain functions of CtIP that are involved in MMEJ.

1.5.3 HR

In contrast to NHEJ and MMEJ faithful DNA repair is achieved through the process of Homologous Recombination (HR). This pathway involves a series of processes which use an undamaged homologous DNA template to lead to successful repair. HR maintains genomic stability through accurate repair of DSBs alongside recovery of replication forks, and contributes to tolerance of DNA damage (Krejci et al., 2012). HR is also critical to maintenance of telomeres and during chromosome segregation in meiosis 1 (where NHEJ cannot act) through the formation of crossovers (Li and Heyer, 2008; San Filippo et al., 2008). Therefore HR must be under tight control as unchecked recombination can lead to gross chromosome rearrangements, loss of heterozygosity and cell death (Kaniecki et al., 2018).

Research into recombination started as early as 1964 when Robin Holliday published the "Holliday model" of recombination at meiosis in fungi, providing the early molecular basis for gene conversion and crossover (Holliday, 1964). This model highlighted the fact that the action occurs post replication with DNA breaks allowing single strands to anneal to the complementary sequence to form a structure linking the two chromatids. This point is known as the Holliday Junction (HJ), which on resolution results in either a crossover or non-crossover product (Holliday, 1964). Further iterative models built on this such as the Meselson-Radding model showing the process was initiated through a single DSB (Meselson and Radding, 1975). Further studies on yeast meiosis in 1983 led to the classic Double Strand Break repair (DSBR) model. According to this model, the DSB ends were resected in the 5'-3' direction with the 3' tail active in strand invasion, on crossover this gave equal amounts of crossover and non-crossover products (Szostak et al., 1983). Studies in *Drosophila melanogaster* revealed the synthesis-dependent strand annealing (SDSA) model for mitotic recombination and a preference for non-crossover events (Nassif et al., 1994).

HR can now be divided into three stages: presynapsis, synapsis and post-synapsis with proteins such as the Rad51 recombinase functioning in all three stages. We now know that the resected 3' ssDNA is rapidly coated by RPA to protect it from nuclease degradation, eliminate secondary DNA structure and activate the ATR pathway. This RPA coating is subsequently exchanged for the Rad51 recombinase through the mediator protein BRCA2 (Davies et al., 2001). This Rad51 nucleoprotein filament can then undergo strand invasion to search for a homologous sequence. During DSBR on base paring with the homologous sequence DNA synthesis at the 3' end results in a displacement loop (D-Loop), while the nascent strand pairs with the other 3' tail. Both 3' ends are extended by DNA synthesis to form a double HJ intermediate. Resolution then forms both crossover and non-crossover products. In contrast, in SDSA only one of the break ends is paired with the homologous DNA and extended, the

DNA is then displaced and annealed to the unpaired end to form non-crossover products (Kowalczykowski, 2015; Morrical, 2016) (figure 1.5.3). These and further biochemical studies using single molecule microscopy have helped define the molecules and pathways that are involved in recombination known today (Kaniecki et al., 2018).

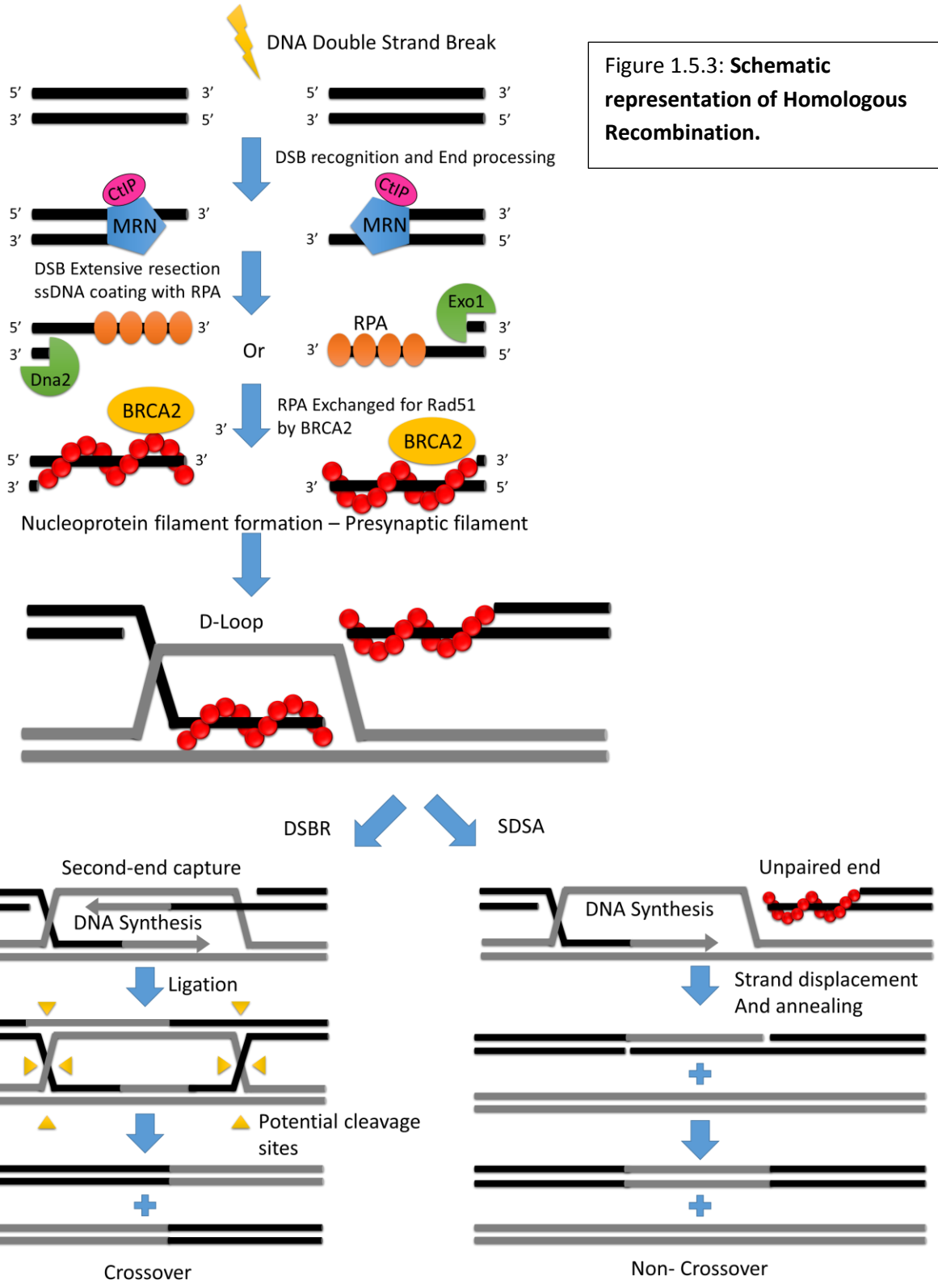


Figure 1.5.3: **Schematic representation of Homologous Recombination.**

1.5.3.1 HR Machinery

In contrast to NHEJ, many of the proteins that are involved in HR were first discovered in the budding yeast *S. cerevisiae.* The classical procedure of mutant isolation by sensitivity to DNA damaging agents and complementation of the mutant phenotype identified the RAD52 epistasis group(Petrini et al., 1997). The structures and functions of the group members were found to be highly conserved with homologues found in the human system such as: Mre11, Rad50, Nbs1, Rad52 and the Rad51 paralogues (Rad51b, Rad51c, Rad51D, XRCC2 and XRCC3) (San Filippo et al., 2008). Many other factors such as RPA and the breast cancer susceptibility proteins BRCA1 & BRCA2 were identified to play crucial roles in HR in vertebrates. Further research in *S. cerevisiae* discovered Sae2, which was later revealed to be the homologue of CtIP (Sartori et al., 2007). Each of these proteins have specific functions within HR with assembly of the presynaptic filament being the rate determining step (San Filippo et al., 2008).

1.5.3.2 MRN, presynaptic filament formation

The initial stage in all DSB repair mechanisms is identification of a DSB. Although originally thought to be the role of ATM, the Mre11, Rad50, Nbs1 (MRN) complex has been attributed to break sensing and recognition for initiation of HR. Its influence however is not limited here, with the MRN complex also contributing to cell cycle control and alongside its critical role during repair of the lesion (Zha et al., 2009).

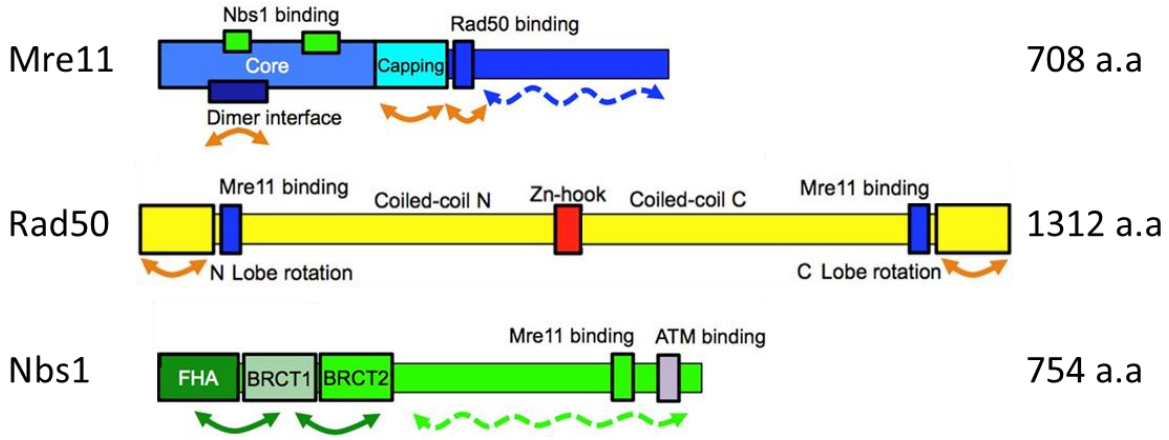


Figure 1.5.3.2-1: **Schematic representation of the MRN complex proteins Mre11, Rad50 and Nbs1.** Flexible regions are shown by solid arrows. Dashed arrows represent disordered segments. Adapted from Lafrance-Vanasse et al., 2015.

The Mre11 subunit of the MRN complex is a highly conserved protein composed of an N-terminal phosphoesterase domain, which possesses manganese-dependent nuclease activity *in vitro*, and two distinct C-terminal DNA binding domains (D'Amours and Jackson, 2002). Isolated Mre11 has been shown to form stable dimers in solution possessing intrinsic DNA binding activity and enzymatic endo- and exo-nuclease activity against both ssDNA and dsDNA substrates, with the ability to act on 5' overhangs, 3' flaps and 3' branches (De Jager et al., 2001a; Lamarche et al., 2010; Mimitou and Symington, 2009). Determination of an archaeal Mre11 dimeric structure led to distinction between endo- and exo- nuclease activities of Mre11, with single point mutations selectively inhibiting activity respectively (Williams et al., 2008). Conversely, the Mre11 subunit lacks the 5' – 3' exonuclease activity thought to be required for generation of 3' tails with its 3' to 5' activity shown to be necessary for the action (Lamarche et al., 2010; Paull and Gellert, 1998).

In vivo, the Mre11 protein exists in an $Mre11_2Rad50_2$ heterotetrameric core complex with each Mre11 molecule binding to a single Rad50. The Rad50 subunit belongs to the conserved family of the structural maintenance of chromosomes (SMC) possessing both structural and sequence homology. Rad50 has been shown to form an antiparallel coiled coil through crystallography studies in archaea folding back at a conserved Cys-X-X-Cys (CXXC) motif. This was shown to form an interlocking hook upon coordination of a Zn^{2+} ion to mediate dimerisation of Rad50 (Hopfner et al., 2002; Lamarche et al., 2010). The sequence homology to the SMC proteins exists at the N and C-termini with Walker A and B motifs present respectively; these stably associate to form an ATP-binding cassette (ABC)-type ATPase domain with DNA binding affinity to dsDNA termini (De Jager et al., 2001b).

Collectively the domains of Rad50 have illuminated the function at the break site in the MR core complex. Deletion of the coils and zinc binding independently have been shown to impair the DNA binding affinity and ATPase activity, showing the coiled-coils link the Rad50 catalytic heads together for action with Mre11 (Deshpande et al., 2014; Hopfner et al., 2002; Lee et al., 2013; Paull and Deshpande, 2014). The ATP-related activities of Rad50 are also essential for its role in DNA repair with mutations in the Walker A motif akin to *Rad50* deletion (Moncalian et al., 2004). Specific changes in the Rad50 coils on ATP binding and hydrolysis change the angle from 90-120° to alter the orientation of Mre11 and accessibility of the nuclease domain to the DNA (Deshpande et al., 2014; Lammens et al., 2011; Lee et al., 2013; Lim et al., 2011; Williams et al., 2011). Analysis of the ATP binding has also shown that the ATP-bound closed state promotes Rad50 DNA binding, tethering of the DNA ends, and ATM activation, but prevents Mre11 nucleolytic activation (Lamarche et al., 2010). With ATP-hydrolysis by both Rad50 active sites necessary for Mre11 activation (Deshpande et al., 2017). Shortening the length of the coils (removing 300 amino acids) has no effect on HR while it does affect telomere maintenance, meiotic DSB formation and surprisingly NHEJ, indicating two roles for the

Rad50 coiled-coil domain in inter- and intra-chromosome bridging (Hohl et al., 2010). In the current model, a dual conformation of Rad50 has been shown. A long straight ridged Rad50 within the MR complex is involved in DNA tethering activities, while a looped Rad50 structure is involved in DSB repair (Lafrance-Vanasse et al., 2015).

The third subunit of the MRN complex, Nbs1 (Xrs2 in yeast), binds to this core complex via Mre11 to form the MRN complex in a 2:2:2 stoichiometry (Lamarche et al., 2010). Nbs1 was originally identified as the mutated protein in Nijmegen breakage syndrome, a rare autosomal-recessive disease with hypersensitivity to IR, immune disorders, microcephaly, growth retardation and cancer predisposition (Chrzanowska et al., 2012). Nbs1 lacks enzymatic activity but can be viewed as both a sensor and an effector due to its Mre11 interaction and ability to recruit other proteins, with links to ATM signalling and modulation of Mre11 nuclease activity binding both through its C-terminal domain (Crown et al., 2013; Lee and Paull, 2004, 2005; Park et al., 2011). At the N-terminal domain of Nbs1 a FHA and two BRCT motifs mediate phospho-dependent protein-protein interactions such as with CtIP (figure 1.5.3.2-2)(Lloyd et al., 2009; Williams et al., 2009), MDC1 (Lloyd et al., 2009b; Spycher et al., 2008), WRN (Kobayashi et al., 2010) and ATR (Olson et al., 2007). From structural determination of Nbs1 in complex with a *S. pombe* Ctp1 (CtIP homologue) peptide long range conformational changes in the N-terminal of Nbs1 were seen having the potential to regulate the BRCT1 & 2 sites to modulate repair pathway choice by CtIP (Williams et al., 2009). Further evidence for the importance of Nbs1 phospho-specific interactions comes from mutation of the FHA domain causing high sensitivity to IR, camptothecin and hydroxyurea known to cause DSBs repaired by MRN highlighting Nbs1's key role within the complex (Lloyd et al., 2009; Williams et al., 2009).

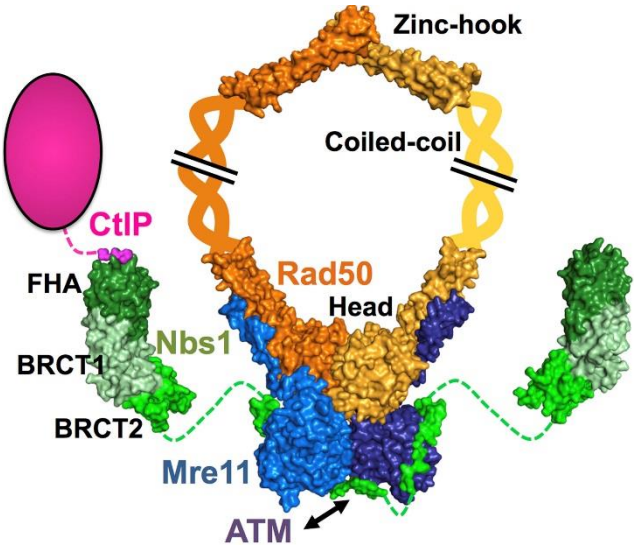


Figure 1.5.3.2-2: **Architecture of the Mre11-Rad50-Nbs1 complex with Nbs1 bound to CtIP.** Where possible space fill models are represented with colours for Nbs1 and Mre11 the same as in **figure 1.5.3.2-1**. Rad50 is represented as a dimer consisting of one monomer in orange, one monomer in yellow. ATM binding is represented by an arrow to Nbs1 and Mre11. Dashed lines represent disorders protein linkers. Adapted from Lafrance-Vanasse et al., 2015.

Following detection of the DSB and DNA binding by Mre11-Rad50, ATM signalling is activated for downstream DSB signalling and necessary checkpoint control. MRN 3' -5' nuclease activity has long been known to be necessary for the subsequent generation of 3' tails. Recently it has been reported that the MRN complex moves 100-200 base pairs away from the end of the DNA after initial detection of the break (Shibata et al., 2014). This was shown to be independent of Rad50 ATPase activity, although it relied on the Rad50 head domain for 1D diffusion along the DNA (Myler et al., 2017). Post diffusion, CtIP/ Sae2 have been shown in nuclease assays for humans and *S. cerevisiae* to stimulate the endonuclease activity of Mre11, cutting the 5' strand 3' to the break site (Anand et al., 2016; Cannavo and Cejka, 2014). Activity in human CtIP is dependent on the conserved threonine 847 (T847) CDK phosphorylation site and interaction with Nbs1 (Anand et al., 2016; Sartori et al., 2007). However, the change in architecture within the MRN complex on phosphorylation of CtIP T847 is yet to be determined. From this point Mre11 5'-3' exonuclease activity resects the DNA strand back towards the DSB end (Cannavo and Cejka, 2014; Garcia et al., 2011; Shibata et al., 2014). While in the opposing direction the BLM helicase along with Exo1 or DNA2 nucleases carry out the long range 5'-3' resection (Gravel et al., 2008; Zhu et al., 2008). This has been shown to be MRN nick-generation dependent (W. Wang et al., 2018) (figure 1.5.3.2-3). MMEJ is inhibited by the long-range resection and binding of RPA to channel the break into HR repair preceding Rad51 presynaptic filament formation by BRCA2.

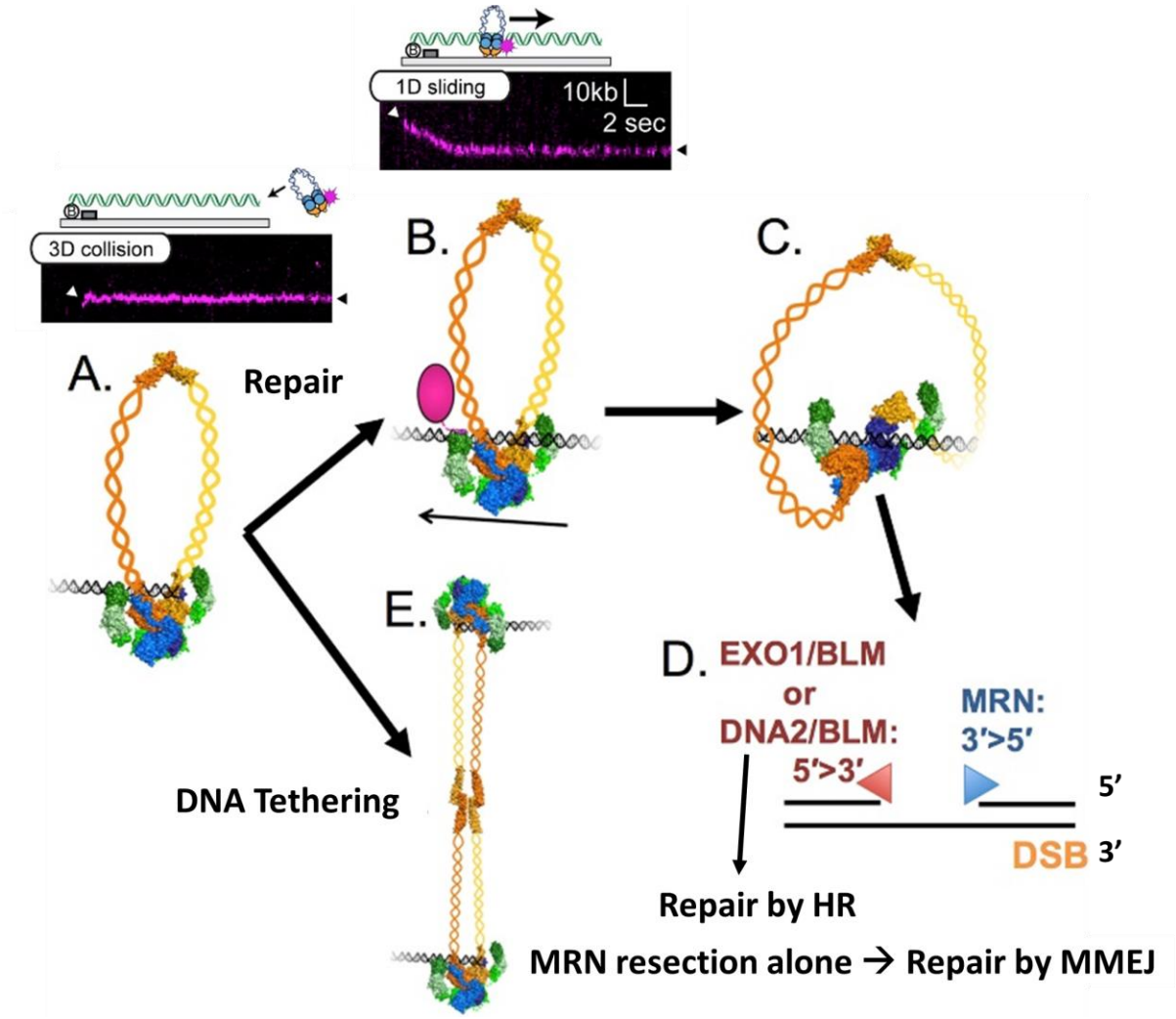


Figure 1.5.3.2-3. **Model of MRN functions in DSB repair.** Colours are the same as in previous figures in this section **A)** Initial detection is by 3D collision with the DSB. **B)** In HR the complex relies on Rad50 head domain in 1D diffusion to move close to the DSB end. **C)** Phosphorylated CtIP bound to Nbs1 induces unknown changes in MRN leading to a complex capable of endonuclease activities to generate a nick 3' to the 5' DSB end. **D)** MRN 3' -5' exonuclease activity resects DNA back to the DSB end, while BLM with Exo1/Dna2 resects in the 5' to 3' direction. **E)** In both NHEJ and HR DNA tethering activates of MRN can be utilised. Adapted from Lafrance-Vanasse et al. 2015, with single molecule fluorescence from Myler et al., 2017.

1.5.4 Regulation of DSB repair pathways

As discussed in Sections 1.4 and 1.5 there is a large array of DNA repair pathways that have evolved to tolerate and repair a variety of different lesions. In the case of DSB repair, regulation between the choice of HR, NHEJ and MMEJ pathways is vital. One method that necessitates HR is the presence of the sister chromatid: limiting HR to G2 and S phase of the cell cycle. In contrast, both NHEJ and MMEJ can operate throughout the cell cycle, however are often mutagenic. Many other methods of regulation exist to control these pathways, for example the availability of HR factors such as Rad51 which has been shown absent in mouse G1 phase (Yamamoto et al., 1996), along with CtIP and BRCA1 which have very low expression at the protein level in G1 (Yu and Baer, 2000).

Although NHEJ is active throughout the cell cycle, it must also be controlled to ensure HR can mediate repair where necessary. Recent studies show that many of the chromosome abnormalities and repair defects in Fanconi Anaemia are a result of inappropriate action of NHEJ in S-phase (Adamo et al., 2010). The 53BP1 and BRCA1 proteins are emerging as pivotal regulators of DSB repair by NHEJ and HR respectively with many genetic studies highlighting their roles in this decision. 53BP1 disruption has been shown to channel breaks into HR, with 53BP1 mutation rescuing lethality of BRCA1 deficiency in mice (Cao et al., 2009), likely by relieving end-resection inhibition at DSBs through RIF1 (Bunting et al., 2010; Jasin and Rothstein, 2013; Xu et al., 2015; Zimmermann et al., 2013). The exact method by which BRCA1 and 53BP1 counteract each other at the biochemical level to choose between NHEJ and HR remains unclear (Chapman et al., 2012). The helicases, BLM and WRN have also been shown to play a role in this decision between canonical NHEJ and HR by modulating the binding of Mre11/CtIP to the DSB (Grabarz et al., 2013; Shamanna et al., 2016). More recently ubiquitinated BLM has been shown to be recruited by MRN to DSB sites to modulate both NHEJ and HR repair in a cell-cycle dependant manner (Tripathi et al., 2018).

Several mechanisms have evolved to reverse the HR reaction through anti-recombinase activities during S-phase to ensure HR is activated at the correct time and location. This can be acted upon at various stages of the process to counteract toxic HR intermediates and to alter the crossover outcome of repair. One method has been shown by yeast a UvrD-like helicase Srs2, which acts to dismantle Rad51-ssDNA filaments to prevent inappropriate HR at active replication forks (Marini and Krejci, 2010). Orthologous to Srs2, PARI has been shown to act *in vitro* by binding and disrupting Rad51 filaments interacting with PCNA which is present at replication forks (Moldovan et al., 2012). Further control of HR post resection and formation of the presynaptic filament exist to control the pathway of recombination and ensure homologous sequence alignment (Heyer et al., 2010).

Despite these mechanisms when NHEJ and HR are simultaneously available the choice between the two pathways remains. Experiments in human cell lines showed that NHEJ is much faster than HR taking place in 30 minutes verses several hours for HR (Mao et al., 2008), highlighting the importance of this decision. In G1 phase the strong DNA binding affinity of Ku outcompetes HR repair factors, whereas experiments overexpressing Ku in G2 show a reduction in the recruitment of Mre11 (Fell and Schild-Poulter, 2015). The decision between the repair pathway can also depend on the complexity of the break, low complexity DSBs are repaired by NHEJ while more complex DSBs with protein adducts or clusters of breaks are repaired by HR (Hada and Georgakilas, 2008; Shibata et al., 2011). Evidence now exists towards a cellular model whereby canonical NHEJ acts first in an attempt to repair the break, with activation of resection when NHEJ cannot repair the break during S-G2 phase (Her and Bunting, 2018; Shahar et al., 2012). The emerging 3' endonuclease CtIP/MRN nick model of DSB repair strongly supports this process, suggesting a pathway for overcoming the strong affinity of Ku at DNA ends. With this, resection by Mre11 has been shown to take place at Ku-bound DNA ends or other DNA-end adducts, requiring ATM phosphorylation of CtIP and Mre11 nuclease activity (Chanut et al., 2016; Myler et al., 2017).

DSB DNA-end resection is a key commitment step for homology directed repair (HDR). Therefore, many converging signals act upon CtIP to control and limit DNA-end resection. One method mentioned previously is that of CDK phosphorylation of CtIP at T847, necessary for its action with MRN and therefore is restricted to S/G2 phases of the cell cycle (Huertas and Jackson, 2009). A further phosphorylation event is that of PLK1 which phosphorylates CtIP at multiple serine residues at the C-terminus (Serine 723 in particular), this has been shown to reduce BLM recruitment and therefore help channel resected DNA into MMEJ (H. Wang et al., 2018). A second regulation focused on CtIP is that of its neddylation by RNF_{111}/UBE_2M, this can affect the choice between NHEJ and HR, and can also regulate the extent of resection and sub-HR-pathway choice (Jimeno et al., 2015). Many other phosphorylation, sumoylation, ubiquitination and isomerization events act on CtIP to regulate its action at the DSB and will be discussed in the next section. How many of these events change the structure of CtIP and elicit physiological actions however are unknown.

1.6 CtIP (CtBP-interacting protein)

CtIP was originally identified in 1998 through its interactions with two proteins: C-terminus binding protein of adenovirus E1A (CtBP) and Retinoblastoma (RB) giving it the names CtBP Interacting Protein (CtIP) and RB binding protein 8 (RBBP8 or Rim) respectively (Fusco et al., 1998; Ute Schaeper et al., 1998). Shortly after its discovery, interactions of CtIP and BRCA1 were also reported (Wong et al., 1998). In the subsequent years the name "CtIP" has been more commonly used, with functions in transcriptional regulation, cell cycle checkpoint control, genome stability, tumour suppression and more recently methods of DSB repair identified (Wu and Lee, 2006).

The human *CtIP* gene is located at chromosome 18q11.2 and encodes an 897 amino acid protein. Many interaction motifs within CtIP have been characterised such as the RB binding motif (LECEE), CtBP binding motif (PLDLS), the MRN binding sites and further direct interaction motifs detailed in the next section of this thesis and shown in figure 1.6-1.

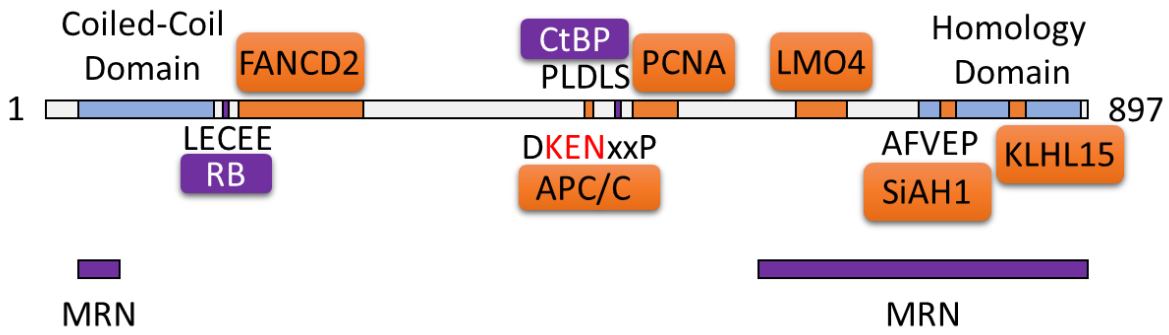


Figure 1.6-1: **Schematic representation of direct interaction motifs on human CtIP.** Interacting partners are shown in orange and purple.

Aside from direct interaction many phosphorylation sites are present in CtIP with some mediating interaction in a phosphorylation-specific manner. Sites on CtIP include ATM phosphorylation at serine 664/745 and threonine 859. Many CDK phosphorylation sites are also present with the importance of serine 327 and T847 well established in CtIP function. Further sites are also detailed in this thesis and in figure 1.6-2.

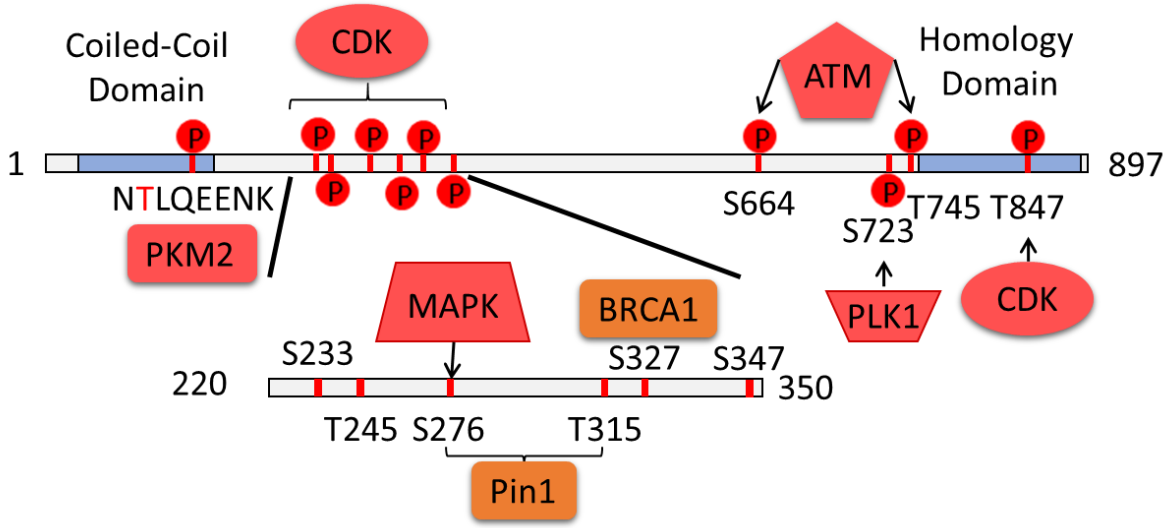


Figure 1.6-2: **Schematic representation of phosphorylation sites on human CtIP.** Kinases are shown in red, binding partners in orange. MAP kinase has also been shown to phosphorylate Serine 276 for PIN1 binding to S276 & S315. Residues S233, T245, S276, T315 and S347 have all been shown to be important in Nbs1 binding. Residue S723 phosphorylated by PLK1 promotes MMEJ

Structurally only short regions have been determined such as the tetramerisation domain from residues 18-52 (Davies et al., 2015) and a short region containing a small beta strand that interacts with LMO4 shown in figure 1.6-3.

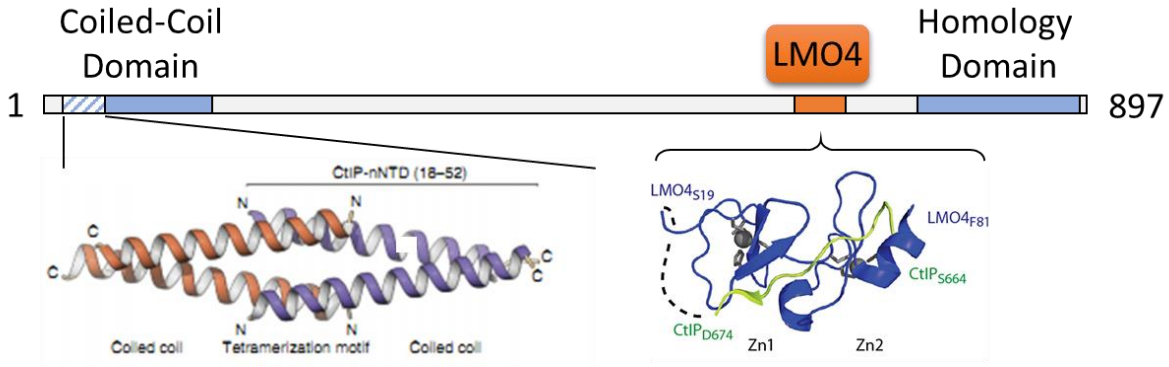


Figure 1.6-3: **Schematic representation of determined CtIP structures within the literature.** Structure images for the CtIP-nNTD and LMO4 were extracted from papers by Davies et al 2015 and Stokes et al 2013.

1.6.1 CtIP and homologous recombination

Although CtIP was originally associated with transcriptional regulation, Sartori *et al.* revealed the role of human CtIP in DNA end-resection prior to ATR signalling pathway activation (Sartori et al., 2007). Here evidence for CtIP's involvement in DSB repair came from 1-hour siRNA depletion of CtIP in U2OS cells, the results showed hypersensitivity to camptothecin and etoposide which cause replication-dependent DSBs. The effects of short CtIP depletion were abrogated by inhibiting replication using aphidicolin. As DSBs generated in S-phase by camptothecin are repaired by HR, it suggests a role for CtIP to promote HR at DSBs. As mentioned previously, on detection of a DSB ATM phosphorylates H2AX with ATR activating Chk1 after ssDNA generation. Camptothecin treatment of cells caused phosphorylation of H2AX (γH2AX) and hyper-phosphorylation of CtIP, however depletion of CtIP did not affect the formation of γH2AX. This indicates that depletion of CtIP still allowed the drugs to cause DSBs and they were still detected by ATM. Conversely, recruitment of ATR and phosphorylation of Chk1 were drastically reduced on CtIP SiRNA treatment indicating CtIP played a role in generation of the ssDNA required for their downstream activation. Analysis of ssDNA tracks showed co-localisation of CtIP with RPA, however on CtIP depletion RPA was absent from the break foci. Anti-BrdU staining for ssDNA confirmed that on depletion of CtIP and camptothecin treatment ssDNA was not formed (Sartori et al., 2007). These data established that CtIP promoted ssDNA formation at a DSB site. Further co-immunoprecipitation experiments showed CtIP could interact with the MRN complex while direct interactions were confirmed using purified regions of CtIP which could retrieve MRN from HeLa cell extracts (Sartori et al., 2007). The MRN CtIP interaction was further confirmed by an independent study by Yuan & Chen (Yuan and Chen, 2009). In addition, the ground-breaking research published by Sartori *et al.* demonstrated activation of the MRN endonuclease activity by CtIP in the presence of magnesium with an increase in activity in the presence of manganese. CtIP alone exhibited no nuclease activity when purified from HEK293 cells (Sartori et al., 2007). Specific involvement of the C-terminal domain in DSB repair through truncation of CtIP was also reported. Here removal of the last 108 CtIP residues inhibited Chk1 phosphorylation and recruitment of RPA without affecting the cell cycle (Sartori et al., 2007).

From this point forward CtIP's role in DSB repair has been further established in generation of ssDNA in complex with MRN. An independent study in 2008 showed interactions of the CtIP N&C-termini with Nbs1 which was significantly strengthened in the presence of Rad50 and Mre11 (Chen et al., 2008). These results were followed in 2009 reproducing CtIP depletion affecting the HR processes downstream of RPA recruitment but not affecting γH2AX formation. Chromatin immunoprecipitation (ChIP) experiments reported by You *et al.* located CtIP to within 100bp of a DSB end with the same results for Nbs1 and Ku70, all three of these required a broken end and would not associate with a

circular plasmid (You et al., 2009). In the same year the CDK sites serine 327 and T847 were reported to promote interaction with BRCA1, deciding between DSB repair pathways and activation of HR respectively (Huertas and Jackson, 2009; Yun and Hiom, 2009). In 2010 the action of CtIP at the DSB site in alignment with its role as a multivalent adapter protein was found to be a controlling factor in initiation of HR repair. Activation of the MRN complex end resection was only seen on integration of checkpoint and cell cycle signals by CtIP (You and Bailis, 2010). Further regulation of CtIP by phosphorylation, sumoylation, ubiquitination and proline isomerization were subsequently discovered in the years to follow and will be covered later in this section.

The discovery of CtIP's involvement in resection is not limited to vertebrates with many functional homologues characterised in other eukaryotic species including *Arabidopsis thaliana* and *Caenorhabditis elegans* through to the *Saccharomyces cerevisiae* homologue Sae2 and the *saccharomyces pombe* homologue Ctp1 (Clerici et al., 2005; Limbo et al., 2007; Uanschou et al., 2007). The C-terminus of CtIP is homologous through to *S. cerevisiae* Sae2. Limited sequence similarity exists within the disordered region of the protein, however homology within the N-terminal tetramerisation domain does exist in Ctp1 and in other species (Andres et al., 2015; Sartori et al., 2007; Uanschou et al., 2007). Analysis of Sae2 null *S. cerevisiae* showed accumulation of un-resected DSBs, reduced synapsis formation and absence of sporulation which resembles Mre11 nuclease-dead mutations (Moreau et al., 1999; Prinz et al., 1997). Ctp1 has also been placed at the DSB junction being required for MRN DNA-end resection and subsequent RPA coating (Limbo et al., 2007). The name "Sae2" originated from the absence of sporulation being recovered in a Spo11 null background (Sae2 is an acronym for Sporulation in the absence of Spo11). Spo11 is known as a type II topoisomerase-like DNA transesterase which is meiotic specific; in a Sae2 null background spo11 remains covalently attached to DNA with Sae2 activity required for the removal (Prinz et al., 1997). Ctp1 has also been reported to play a role in removal of topoisomerases from the 5' and 3' DSB break ends with Rad32 *(S. Pombe* Mre11). Here Ctp1 preferably removes 5' linked Top2 while protecting 3' linked Top1 suggesting a role in 3' end protection (Hartsuiker et al., 2009).

Alongside the critical role in DSB repair evidence also exists that CtIP plays a crucial role in ICR upstream of a DSB, interacting with the FA protein FANCD2. Residues 166-273 of CtIP have been shown to be sufficient for the FANCD2-CtIP interaction recruiting it to the damaged chromatin. Here, the interaction tethers CtIP at the stalled or collapsed replication fork to encourage restart by HR. This interaction supresses the action of NHEJ at the fork which is a less favourable pathway for fork restart. Studies in CtIP depleted cells with aphidicolin treatment (to stall replication forks) showed that CtIP not only has roles in replication fork restart but also in repression of new origin firing (Duquette et al., 2012; Murina et al., 2014; Unno et al., 2014; Yeo et al., 2014).

1.6.2 Regulation of CtIP

With the great power of CtIP acting in a multitude of pathways comes a great responsibility to ensure it is under tight regulation. Abundance of CtIP protein is controlled through its promotor activity and at the post-transcriptional level; specific functions of CtIP are also regulated by modifications such as phosphorylation and sumoylation, and by binding different interaction partners. CtIP protein levels are limited in G1 phase, however expression is up-regulated during late G1 phase. Consistent with its role in transcriptional regulation CtIP has been shown to control its own gene expression to maintain protein levels throughout S-phase (Wu and Lee, 2006).

During S and G2 phase of the cell cycle CtIP can be phosphorylated by CDK's, with 12 potential sites in human CtIP (Ferretti et al., 2013); remarkably phosphorylation of a single CDK site, T847, is required to promote resection. Mutation studies of the highly conserved T847 showed impaired RPA recruitment when substituting to alanine which was not the case when mutating to a phosphomimetic glutamic acid residue. In both cases, however, DSB repair efficiency was affected causing hypersensitivity to DSB-generating agents suggesting a fine level of control through the CDK site (Huertas and Jackson, 2009). Similar studies using Sae2, showed residue Serine 267 was equivalent and also required for DNA resection with a phosphomimetic allowing resection in the absence of CDK (Huertas et al., 2008). More recently phosphorylation of CtIP at T847 has been shown to be the key step in activation of MRN endonuclease activity to initiate DNA cleavage on the 5' strand. This then allows activation of MRN's 3'-5' exonuclease activity and generation of a 3' overhang in the presence of a protein block, the same has also been seen for Sae2 however without the need for phosphorylation (Anand et al., 2016; Cannavo and Cejka, 2014).

The actions of the T847 site have also been shown to be independent of the interaction with BRCA1, which interacts through a further CDK site at serine 327 (Polato et al., 2014). The role of the serine 327 CDK site for this interaction is still under debate, however evidence shows that the phosphorylation directs cells towards HR rather than MMEJ showing a role in switching the repair pathway for DSBs (Polato et al., 2014; Yun and Hiom, 2009). CtIP's role in MMEJ and the switch between the pathways by BRCA1 could present an effective target for breast cancer treatment that display MMEJ-dependent chromosomal rearrangements (Reczek et al., 2016). More recently advances in single molecule studies of resection tracks have given a better understanding of the BRCA1-CtIP interaction in the context of DNA end resection. Cruz-Garcia *et al.* showed that when using a serine 327 to alanine mutant protein the rate of resection was reduced with conclusions from their paper indicating a role of BRCA1 to modulate the speed of resection (Cruz-García et al., 2014). Phosphorylation of CtIP can also be mediated by other kinases such as PKM2 (T126), this has been

reported to increase DSB repair efficiency and improve resistance to genotoxic treatments (Sizemore et al., 2018). Phosphorylation by PLK1 at the C-terminus (S723), as well as DNA damage-inducible PLK3 phosphorylation at T847 and S327 in G1 phase, has been shown to regulate CtIP and direct the pathway choice towards MMEJ (Barton et al., 2014; H. Wang et al., 2018) highlighting the role of CtIP as an orchestrator of cellular events.

CDK-mediated phosphorylation of CtIP has been shown to be a prerequisite for ATM-mediated phosphorylation of CtIP and subsequent recruitment of BLM and Exo1 to the DSB site (Wang et al., 2013). After DNA damage CtIP is phosphorylated by ATM on serine residues 664 and 745, however the function of this is unclear. One possibility is that this causes a conformational change that exposes a DNA binding region in the middle of CtIP. Live-cell imaging of laser-line micro irradiation of U2OS cells showed ATM kinase activity was necessary for recruitment of a CtIP minimal damage recruitment motif (509-557) (You et al., 2009). Conversely, the region 505-546 has also been shown to interact with PCNA, which binds to DNA and initiates MMR as discussed previously (Gu and Chen, 2014). As the first study used camptothecin, where the damage caused is repaired by MMR, the interaction of this region to DNA is more likely to be through PCNA (Fedier et al., 2001). ATM phosphorylation is still potentially necessary to expose this region as its requirement in topotecan (camptothecin analogue) induced replication associated DSB's has been shown (Köcher et al., 2013).

Sumoylation has been shown for Sae2 to increase solubility of the protein as a further control mechanism; without this modification the repair function of Sae2 is impaired (Sarangi et al., 2015). In human CtIP, sumoylation occurs on lysine 896 by the SUMO E3 ligase CBX4. Depletion studies of CBX4 confirmed impaired DNA end processing without the sumoylation, which blocked HR and decreased genome stability (Soria-Bretones et al., 2017). A further layer of CtIP regulation is that of acetylation (Rajendran et al., 2013). Under unperturbed conditions CtIP was shown to be acetylated which was abrogated on DNA damage. CtIP Acetylation was shown to inhibit DNA end resection, with activation of end resection on deacetylation, however the publication has now been retracted (Kaidi et al., 2010).

Controlling CtIP by degradation is another mechanism to regulate its activity through protein levels. The Sartori group has identified a pathway proposed to fine-tune DNA end resection using the Cullin3-KLHL15 ubiquitin ligase and the Prolyl Isomerase PIN1. In the case of PIN1 the interaction is with two CDK sites serine 276 and threonine 315 which on phosphorylation promote CtIP isomerization. This action then leads to ubiquitination and proteasome degradation. In the quest to find the protein that ubiquitilates CtIP in this pathway interactions with the Cullin3-KLHL15 complex were discovered. Here the KLHL15 protein binds to a short region in the C-terminus of CtIP relying on the side chain aromatic ring of tyrosine 842 rather than its phosphorylation. This interaction which is stabilised by action of

PIN1 then causes polyubiquitination of CtIP by Cullin3-KLHL15 for subsequent degradation by the proteasome (Ferretti et al., 2016; Steger et al., 2013). Further control of CtIP by protein degradation is through its conserved patch between amino acids 464 and 472 which contain the recently reported consensus KEN box sequence D/N-K-E-N-x-x-P. This targets CtIP for degradation through the anaphase-promoting complex/cyclosome-Cdh1 (APC/C(Cdh1)) with studies showing this occurs during mitotic exit and during control of DSB repair (Lafranchi et al., 2014). A further E3 ligase SiAH1 has been shown to bind CtIP, SiAH1 is induced by p21 and p53 during apoptosis and tumour suppression. The association of SiAH1 with CtIP has been shown to cause CtIP degradation through the ubiquitin proteasome pathway with SiAH1 having a critical role in breast and prostate cancers (Germani et al., 2003; Knauer et al., 2015).

It is likely that convergence of different signals on CtIP changes the outcome of its interactions and the structure of CtIP. Evidence to support this comes from research by the Charpentier group working on Cas9 to encourage gene integration by HR using the CRISPR Cas9 system. Here they discovered a preference to HDR when Cas9 was fused to CtIP. Further study into the domains of CtIP showed a region spanning 1-296 sufficient for HDR activation. Within this region 3 out of the 6 CDK sites involved in Nbs1 binding were shown to be necessary for the enhancement (serine 233, threonine 245, serine 276) (Wang et al., 2013). In contrast to a full-length fusion, CtIP residues 1-296 fusion to Cas9 gave a different pattern of insertion deletions showing that the further domains of CtIP change the function (Charpentier et al., 2018). Two key points of interest come from these experiments. Firstly, not all the Nsb1 phosphorylation sites were necessary, indicating distinctive regulation and effects of each Nbs1 binding site. Secondly, the C-terminal half of CtIP containing T847 was dispensable for the actions, meaning CtIP must affect MRN without the proposed T847 endonuclease activity activation of MRN (Anand et al., 2016).

Although this fusion and function is manmade, it brings about many questions as to the known and unknown functions of domains within CtIP and what they mean within a wider cellular context. Further study into the structure and action of the CtIP domains and the modifications brought about by the CDK phosphorylation events are therefore necessary, especially in the context of its interaction with MRN.

1.6.3 Transcriptional regulation by CtIP

Further roles in transcription have been established since identification of CtIP through its CtBP interaction. As mentioned previously, CtIP regulates its own transcription. During late G1 phase CtIP is recruited to its own promoter to associate with RB to release E2F, CtIP then remains at the promotor

site to activate its own transcription by binding to the transcription machinery TFIIB and TATA binding protein (TBP) (Koipally and Georgopoulos, 2002). This action of CtIP also activates transcription of other E2F controlled genes such as Cyclin D1 which are essential to enable cell cycle progression from G1 to S phase (Liu and Lee, 2006; Wu and Lee, 2006). More recently, sustained E2F-dependant transcription with increased cellular levels of CtIP has been shown to be a key mechanism in preventing replication-stress-induced DNA damage (Bertoli et al., 2016).

The interaction of CtIP and BRCA1 also facilitates transcriptional regulation. For example, both CtIP and BRCA1 are required to form a complex with LMO4 which represses BRCA1-mediated transcriptional activation through inhibition of BRCA1 binding to other proteins (Sum et al., 2002). CtIP also regulates gene transcription through its binding to other transcription factors. For instance, binding to CtBP can in turn recruit further proteins involved in promoter targeting and chromatin remodelling (U. Schaeper et al., 1998). The same is also true for the CtIP-RB interaction connecting CtIP through RB to the Krupple zinc finger protein Ikaros for transcription repression (Koipally and Georgopoulos, 2002).

On DNA damage, CtIP has also been shown to play a transcriptional role initiating the production of CDK inhibitor P21, with CtIP binding to the p21 promotor to drive expression (Liu et al., 2014). At the same time, DNA damage activation of ATM leads to phosphorylation of CtIP on serine 664 and 745, it is thought that this can also regulate the expression of ATM as well as help activate cell checkpoint control (Liu and Lee, 2005). The phosphorylation of CtIP by ATM has also been shown to cause disassociation of BRCA1 from CtIP which allows transcription of the GADD45 gene upon IR (Li et al., 2000), this activates p53 through the MAP Kinase pathway for cell cycle control (Salvador et al., 2013).

The Notch signalling pathway can determine cell fate of various cell lineages. Binding of ligand to the receptor leads to release of the Notch intracellular domain which can activate transcription, however, without ligand the Notch controlled genes are supressed by the RBP-Jk/SHARP corepressor complex. Research by Ozwald et al. identified interactions between SHARP C-terminal and CtIP which were sufficient to supress CtIP mediated transcription prior to Notch signalling. The Notch target gene *Hey1* was also shown to be regulated by CtIP/CtBP with CtBP-depressed embryonic fibroblasts showing a strong reduction in *Hey1* (Oswald et al., 2005). Deficiency in *Hey1* known to cause death after embryonic day 9.5 showing the crucial role of CtIP transcriptional control in development (Fischer et al., 2004).

Further evidence exists for CtIP's involvement repair of DSBs in the context of topoisomerase and transcription. GAT3, a transcription factor critical for mammary gland development, has been shown to bind to the CtIP promotor to trigger transcription when topoisomerase II is stalled by etoposide and

DSB are formed (Zhang et al., 2017). CtIP has also been shown to interact with TopBP1 in a damage dependant manner while interacting the Nbs1 bringing the MRN complex to topoisomerase DSBs for repair (Ramírez-Lugo et al., 2011).

1.6.4 CtIP and And-1

The acidic nucleoplasmic DNA-binding protein 1 (And-1) has a well-established role at the replication fork to couple the CMG-helicase and DNA polymerase alpha ensuring fast replication. And-1 has been shown to have many other functions here at the replication fork including its role in the replication pausing complex, recruitment of polymerase alpha to chromatin, activation of Chk1 during replication stress and more recently a function in fork protection preventing excessive spans of ssDNA which are converted to DSBs (Abe et al., 2018; Errico and Costanzo, 2010; Hao et al., 2015; Yoshizawa-Sugata and Masai, 2009; Zhu et al., 2007). Depletion of And-1 has been reported incompatible with replication, causing cells to accumulate in G2 phase due to activation of DNA damage checkpoint control. Within the fork protection aspect of And-1 it was shown that loss of And-1 allows Mre11 to resect DNA to produce long ssDNA stretches. Removal of the And-1 WD40 domain had the same effect as knockdown of And-1 indicating this region to be important for the function (Abe et al., 2018).

Since And-1 and Mre11 are connected in the fork protection pathway is it not surprising that interactions with CtIP have been reported. Identification through a co-immunoprecipitation interaction study has shown that CtIP and And-1 can interact through its C-terminal and a region spanning from residues 850-1050 of And-1. GFP-tagged And-1 immunofluorescence of U2OS cells showed co-localisation to γH2AX sites of DSB prior to CtIP. And-1 knockdown showed a dramatic inhibition of CtIP localisation and reduced subsequent Rad51 foci formation. Measuring HR efficiency on siRNA knockdown of endogenous And-1, and addition of siRNA resistant And-1 that lacked the CtIP interaction patch, failed to restore HR with hypersensitivity to DNA damaging agents. Addition of siRNA resistant wild type And-1 did however recapitulate resection. The results published by Chen et al. showed an important role for the newly discovered And-1 CtIP interaction in the HR pathway, with cells sensitized to DNA damage and replication stress. An increased proportion of chromosome abrogations were also reported suggesting the interaction plays a role in maintaining genome integrity (Chen et al., 2017).

1.6.5 The endonuclease activity of CtIP

Studies on Sae2 showed that it possesses endonuclease activity independent of Mre11 (Lengsfeld et al., 2007). However, a recognisable nuclease active site is absent from the amino acid sequence with other groups unable to show nuclease activity (Cannavo et al., 2018). Unlike the MRX nuclease (yeast homologue to MRN) which requires manganese, Sae2 can operate with either manganese or magnesium to digest DNA. This activity is strongly activated by adjacent DNA hairpin structures which have been shown to be cleaved cooperatively between Sae2 and MRX (Lengsfeld et al., 2007). In contrast, although many other functions are comparable between Sae2 and other homologues, in Ctp1 no intrinsic nuclease activity has been seen (Andres et al., 2015).

There has long been debate within the CtIP field regarding an intrinsic enzymatic nuclease activity. No known nuclease fold can be detected from the human amino acid sequence or through structural predictions. In 2007 Sartori *et al.* assessed the nuclease activity of HEK293 expressed CtIP with no activity present (Sartori et al., 2007). However, the T. Paull group, through various experiments using insect cell expressed CtIP showed nuclease activity. A region between residues 165 and 347 was shown to be important with asparagine 289 and histidine 290 necessary, with claimed homology to Mre11 H129N. The previously mentioned “minimum DNA interaction region”, which was later shown to also bind to PCNA, was also required for the nuclease activity (Makharashvili et al., 2014). A further group also expressing CtIP from insect cells reported evidence of CtIP’s nuclease activity, however both groups confirmed that the action is outside of canonical HR (Makharashvili et al., 2014; Wang et al., 2014). In contrast, expression of CtIP in human cells has not been reported to have nuclease activity (Anand et al., 2016; Liao et al., 2016). Although CtIP nuclease activity cannot be ruled out it remains to be solidified. One possibility is that the different purification methods yield different post translational modifications or co-purify different binding partners from the host expression system. These events could enforce an unknown enzymatic conformation within CtIP.

1.6.6 Oligomerisation and DNA binding of CtIP

Although sequence homology is more limited at the N-terminal domain, oligomerisation of CtIP does exist throughout various orthologues of CtIP. From analytical ultracentrifugation experiments Sae2 has been reported to form several multimeric complexes with a dimeric form dominating. DNA binding of Sae2 has also been seen through electrophoretic mobility shift assays forming a stable complex. Consequently, removal of the N-terminal region resulted in monomeric protein which lacked nuclease activity both with and without MRX and failed to bind DNA. More recently two domains were identified at the N-terminus that are required for oligomerisation and confirmed the functional importance (H. S. Kim et al., 2008). Original discoveries of CtIP identified the N-terminal to be the

region for dimerization through a coiled-coil domain, this dimerisation was reported to stabilise the interaction with RB (Dubin et al., 2004; Stokes et al., 2007). More recently Ctp1 and CtIP have been shown to form a dimer of dimers orientation to generate a tetrameric species. Crystal structures for the start of the domain were published showing a head to head orientation of two parallel coiled-coils. Removal of the N-terminal region in both cases when measuring HR showed the domain to be critical for DSB repair (Andres et al., 2015; Davies et al., 2015). DNA binding activities of the C-terminal domain were also reported. For Ctp1 the tetramerisation made bridging of two DNA strands possible through the DNA binding at the opposing C-termini, weak DNA binding was also reported for the N-termini of Ctp1. For human CtIP DNA binding was only reported for the C-terminus of the protein binding to dsDNA (Andres et al., 2015; Davies et al., 2015).

Within the oligomerisation of CtIP, the extent to which the dimerisation extends remains unclear. Early reports indicated a secondary coiled-coil at the C-terminus of CtIP however with low probability (Dubin et al., 2004; Fusco et al., 1998). Interestingly for the interaction of CtIP and LMO4 the dimeric region of CtIP is required, showing that even ~500 amino acids away from the N-terminal coiled-coil dimerisation to an extent must still be occurring for the interaction with LMO4 (Stokes et al., 2013).

1.6.7 Seckel and Jawad syndromes

Complete deficiency of CtIP in humans has not been seen. Inactivation of CtIP through deletion studies or removal of the T847 CDK site in mice leads to early embryonic lethality. This is due to inability to overcome the G_1 to S phase checkpoint (Chinnadurai, 2006), with haploid insufficiency leading to tumorigenesis and cancer progression in mice (Chen et al., 2005; Polato et al., 2014).

However, two related disorders - Seckel and Jawad syndromes - have been shown to be caused by a frameshift or splicing mutation respectively in CtIP. Despite the recessive inheritance of the disorders, these mutations carry a dominant negative manifestation at the cellular level. Although normal expression levels of CtIP protein are seen, only low levels of truncated CtIP mRNA are present, while the ratio of mutant to wild type protein determines the strength of the disorder phenotype. Expression above the threshold to impede DNA DSB end resection and subsequent ATR signalling causes significant clinical symptoms with an early onset, such as mental retardation, dwarfism, microcephaly, skeletal abnormalities, and chromosome instability (Qvist et al., 2011). Further characterisation of patient-derived cell lines has shown hypersensitivity to DNA damage and a lower threshold for apoptosis. These phenotypes have been attributed to a truncated CtIP produced by the mutations. Notably in cases of Seckel syndrome only the last 152 amino acids are removed highlighting the importance of the C-terminal domain in CtIP function (Qvist et al., 2011).

The magnitude of the phenotypes from such a small discrepancy in a subset of CtIP protein highlights the importance of its known and unknown functions both In DNA repair and in a larger cellular context. Studies to determine the structural and biochemical characteristics of CtIP will gain valuable insight into the protein and help elucidate its role within the complex interacting pathway of the cell.

1.7 Aims and objectives

CtIP has been demonstrated to be a multivalent adaptor protein and an apical point of signal transduction within cellular processes. So far this includes an involvement in transcription, translation, checkpoint activation and now a firm role in DNA DSB repair, with intimate links between the various functions of CtIP. Considering the multiple functions of CtIP and its interacting partners it is expected the structure of CtIP plays a crucial role.

The main aim of this project is to further our understanding of CtIP's function in maintaining genomic stability. To this end, the objectives of the work summarised in this thesis are to purify and characterise the full-length CtIP protein and the individual N- and C-terminal domains using biochemical and biophysical methods. For the full-length CtIP protein the goal is to generate sufficient amounts of protein for biophysical analysis of the reported oligomeric state and DNA binding abilities, alongside structural studies using Cryo EM. The C-terminal domain objective is to confirm the DNA and zinc binding capacity along with interactions with And1, while probing the region adjacent for its oligomeric state. The objective of N-terminal domain work is to determine the crystal structure of the coiled-coil region to assess the impact of the domain on CtIP's architecture.

Chapter 2, Methods

2.1 Molecular Cloning for recombinant protein production

2.1.1 Sequence analysis

Protein sequences of the C-terminus of CtIP human and its eukaryotic orthologues were extracted from the NCBI protein database. The sequences were aligned by ClustalW multiple sequence alignment algorithm (Larkin et al., 2007), and displayed using Jalview (Clamp et al., 2004; Waterhouse et al., 2009). A prediction of coiled–coil structure within CtIP was generated using COILS (Lupas et al., 1991), MARCOILS (Delorenzi and Speed, 2002) and Paircoil2 (McDonnell et al., 2006).

2.1.2 Cloning

For amplification of the N-terminal domain (NTD) constructs PCR was performed on either CtIP image clone 4830693 (Source Bioscience) or a synthesised gene fragment (G-block IDT) generated using the same sequence. CtIP residues 1-897 ($CtIP_{FL}$) for pACEBac1 or pACEMam1 constructs were amplified from cDNA (mRNA variant X3) which was kindly donated to the lab by Tanya Paull. Constructs spanning CtIP residues 694-775 (dCTD) and 694-897 (CTDex) were selected through coiled-coil prediction software and amplified from cDNA (mRNA variant X3) with all primers in Appendix 2. PCR strategy used consisted of initial denaturation at 98°C for 3 minutes, followed by 25-35 cycles of: denaturation at 98°C for 30 seconds, annealing at the appropriate temperature for 30 seconds, followed by extension at 72°C for 30 seconds per 1 kb (minimum 30 seconds) using Phusion® polymerase (Thermo Scientific).

PCR products were cleaned by a PCR purification kit (Thermo Scientific) using manufacturer's instructions and digested with appropriate restriction endonuclease (Appendix 2) at 37°C for 1 to 2 hours. Digested DNA was purified using a DNA Clean and concentrator (Zymogen D4003) kit prior to ligation. Alongside this, the vector (pMAT11 – Bacterial expression with N-terminal His-Maltose Binding Protein (MBP tag), pACEBac1 – Baculovirus expression, pACEMam1 – HEK293F expression) was digested with the same enzymes, dephosphorylated with Antarctic phosphatase (NEB) for 45 minutes followed by 2 minutes 80°C heat kill, and either gel-extracted using Thermo Scientific gel extraction kit (Cat No. K0691) or purified using a DNA Clean and concentrator (Zymogen D4003) kit to give purified cut vector. DNA concentrations were analysed using a Nanodrop™ spectrophotometer at 260 nm. Quick Ligation™ kit (NEB) was used to ligate the cut PCR product into the cut vector at a 3:1 mass ratio with 50 ng of vector used. Ligated plasmid was transformed immediately.

Heat-shock bacterial transformation was performed on Dh5α (NEB) or Dh10β (NEB) bacterial cell lines. 1 to 2 µL of the Quick Ligation reaction was added to 25 or 50 µL of cells and incubated for 30 minutes on ice. Cells were heat shocked at 42°C for 45 seconds followed by 5 minutes of recovery at 4°C. 300-900 µL of SOC medium was added and cells were transferred to a 10 mL culture tube for incubation with orbital shaking at 37°C for 1 to 2 hours. 300 µL of the cell culture was plated onto LB agar plates containing the appropriate antibiotics (pMAT11 – Ampicillin, pACEBac & pACEMam – Gentamicin) and incubated at 37°C overnight.

Colonies were isolated and grown in liquid media for plasmid extraction. Restriction digest was performed using the original insertion primers to confirm incorporation of the insert. For $CtIP_{FL}$ *Xho*I and *Sal*I were used to confirm insertion. Positive clones were sequence-confirmed using the appropriate plasmid-specific primers along with internal primers where required (Appendix 2) by the Department of Biochemistry DNA Sequencing Facility, University of Cambridge.

For addition of an N-terminal Twin Strep tag (appendix figure 7.7) the sequence-confirmed $CtIP_{FL}$ construct was digested with the *Sal*I restriction endonuclease. The amplified Twin-Strep sequence from a synthesised gene fragment was also digested with *Sal*I, and ligation was performed with a 3:1 insert to vector mass ratio. Positive clones were identified by digestion with *Bst*BI (NEB) (appendix 7.7). Positive clones were then sequenced to confirm the Twin Strep tag was in the correct frame orientation with $CtIP_{FL}$.

2.1.3 Mutagenesis

The Cluster mutations and L27E were introduced by the QuickChange® site directed mutagenesis (Strategene) method using manufactures guidelines and primers detailed in Appendix 2. Zinc binding mutations C89A/C92A were introduced to a synthesised gene fragment and sub cloned into the pMAT11 expression vector.

CtIP residues 769-897 (CtIP-CTD) point mutations and combination mutations were generated by overlap extension PCR (figure 2.1.3). Two initial PCR reactions were performed using template, mutation primer pairs at an annealing temperature of 55°C (figure 2.1.3a – part 1). These two PCR products were purified by gel electrophoresis. For the second stage, 5 PCR cycles were performed with these two extracted DNA products with a 55°C annealing temperature. Only end primers were then added and a further 28 cycles of PCR performed at a 64°C annealing temperature (figure 2.1.3a part 2). For combination mutations sections of CtIP-CTD were amplified using the same technique (figure 2.1.3b part 1). Second round PCR was used to anneal together and amplify the DNA to generate first round combination mutant PCR products (figure 2.1.3b part 2). For mutations in close proximity, alternate primers were used to gain multiple mutations from a single primer. Addition of non-mutated

CtIP-CTD DNA or different combinations of mutated DNA template then allowed generation of a multitude of mutation products through successive rounds of PCR (figure 2.1.3b – part 3).

These were then sub-cloned into the pMAT11 expression vector to generate His-MBP fusion proteins for purification. Other PCR constraints followed manufacturer's protocol. All constructs were sequence-confirmed by the DNA Sequencing Facility, Department of Biochemistry, University of Cambridge.

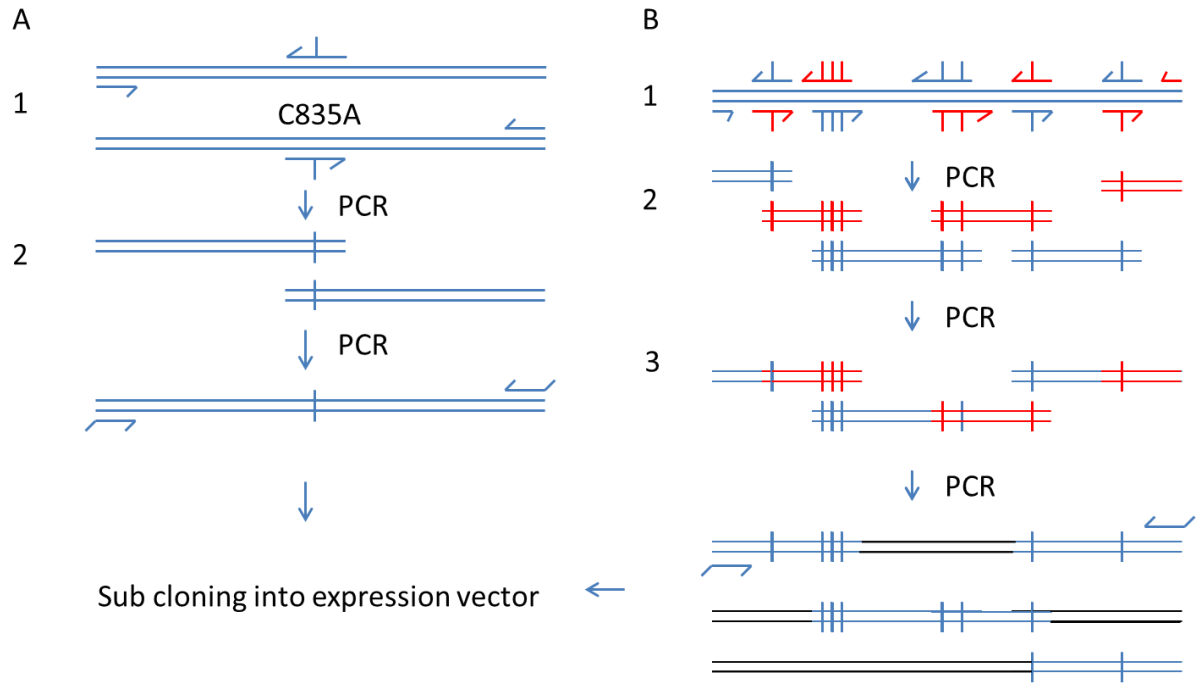


Figure 2.1.3: **Overlap extension PCR method.** Arrows indicate primers; perpendicular lines indicate mutations within primer. Fragments were gel purified after PCR. **A)** Process for single mutants. **B)** Process for combination mutants, with black lines indicating amplified wild type DNA. Multiple rounds of the same protocol were performed consecutively, each contained two DNA fragments at a time until the full length was restored to form the multiple mutations constructs. See text for details.

2.2 Recombinant protein production in *E. coli*

All buffers were filtered and degassed using a 0.22 µm membrane (Triple Red FPE214500).

E. coli Rosetta™ 2 (DE3) cells were transformed using the heat-shock protocol mentioned previously. Antibiotic resistance was used as selection marker for the sequence confirmed construct and addition of chloramphenicol for the T-rare Rosetta™ 2 plasmid. Transformed cells were grown on LB-agar plates until a lawn of colonies was present (16-24 hours) before starting a pre-culture of 10-40 mL in 2XYT grown at 37°C 200 RPM with the same antibiotics. The pre-cultures were then inoculated into between 2 to 8 litres of 2XYT pre- warmed to 37°C with the same antibiotic selection. Cells were grown at 37°C 200PRM with the optical density at 600 nm (OD_{600}) being measured regularly. For growth of cells with zinc-supplemented media when OD_{600} reached 0.4, zinc acetate was added to a final concentration of 250 µM. When cells reached between an OD_{600} of between 0.8 and 1.0, IPTG (Generon) was added to a final concentration of 0.5 mM to induce protein production, at this point the temperature was reduced to 20°C for overnight expression. The cells were grown for 16 h before they were harvested by centrifugation at 4,000 RCF (relative centrifugal force) for 10 minutes at 20°C. Cell pellets were suspended in 24 mL of standard buffer (20 mM Tris pH 8, 300 mM NaCl) with SigmaFAST™ EDTA-Free protease inhibitors per 2 litres of culture before flash-freezing in liquid nitrogen.

2.2.1 *E. coli* cell lysis

The resuspended cell pellets were thawed and either lysed by sonication or using a high-pressure homogenizer (Avestin Inc EmulsiFlex C5). For sonication, cells were lysed within their individual 50 mL tubes on ice for a net time of 3.0 minutes, with a 5 second pulse every 10 second sonication interval (Sonics Vibra-ell™.) When using a high-pressure homogeniser (EmulsiFlex), cells were passed through a total of 3-4 times, maintaining lysis pressure between 5,000 and 10,000 PSI (pound per $inch^2$). Multiple tubes were combined prior to and re-distributed on exit from the EmulsiFlex.

2.2.2 Protein purification of N-terminal His-MBP tagged CtIP NTD

All protein purification stages were carried out at 4°C unless stated otherwise.

A resuspended pellet from 2 litres of *E. coli* Rosetta™ 2 (DE3) cell expression of a NTD construct were thawed and lysed by sonication, followed by clarification by 45,000 RCF for 1 hour at 10°C and subsequent 5.0 µM filtration (Minisart 17594-K). Filtered supernatant was then bound to a 2ml Ni-NTA column (Qiagen 30230) by gravity flow dropwise, washed with 20 mL of wash buffer (standard buffer with 10 mM imidazole) before elution with 9 mL of the same buffer with 200 mM imidazole. Ni-NTA elution fractions containing CtIP-NTD were bound to a 5 mL Amylose (NEB) column by gravity flow dropwise and washed with 10 mL of standard buffer. For removal of the His-MBP tag the resin

was re-suspended in 10 mL of standard buffer inside the chromatography column with TEV protease and 0.5 mM TCEP added. The slurry was rolled at 4°C overnight. Amylose resin was incubated without rolling for 5 minutes prior to collection of the flow through. Five 5 mL washes were performed with the standard buffer. The flow through and first wash fractions were pooled, and 10 mM imidazole added for a 1 mL Ni-NTA column recapture of contaminants. Ni-NTA flow through was collected and concentrated to 1 mL using a 10,000 MWCO concentrator (VivaSpin VS2002). For final polishing a Superdex 200 16/60 Size Exclusion Chromatography (SEC) column was equilibrated into standard buffer, protein applied, and fractions collected. Fractions corresponding to the CtIP-NTD construct were selected by SDS-PAGE, pooled and concentrated using a fresh 10,000Da MWCO concentrator until the protein concentration was above 10 mg/ mL. Concentrated protein was centrifuged at 4°C, 14,000 RCF for 15 minutes before the concentration was measured by absorbance at 280 nm. For constructs inclusive of residues 28-145, protein concentration was determined through comparison to an amino acid analysis standard curve. Protein was aliquoted, and flash frozen in 10/20 µL shots and stored at -80°C for long term storage.

2.2.3 Purification of N-terminal His-MBP, C-terminal Strep tagged NTD and dCTD constructs

For each purification 2 litres of cells were thawed and lysed by high-pressure homogenisation (EmulsiFlex) to disrupt the cell membrane. Lysate was clarified by centrifugation at 45,000 RCF for 1 hour at 10°C, followed by filtration through a 5.0 µm membrane (Minisart 17594-K). For initial capture filtered supernatant was bound to a 2 mL Ni-NTA column by gravity flow dropwise, washed with 20 mL of wash buffer (standard buffer with 10 mM imidazole) before elution with 9 mL of standard buffer with 200 mM imidazole. Protein was dialysed overnight into the standard buffer + 2 mM DTT (Sigma) with addition of TEV were necessary to remove the His-MBP tag. For MBP-fused protein addition of TEV was omitted. To separate the cleaved protein from the His-MBP tag a 2 mL Strep-Tactin™ Superflow® (Strep-Tactin) column was prepared and equilibrated into the standard buffer. The dialysed and cleaved protein was then added to the Step-Tactin resin (IBA 2-1206-025) column and incubated for 5-10 minutes without flow. Using a flow restrictor, the gravity flow of the column was restricted to 0.1-0.2 mL/min to allow the protein to bind. The column was then washed with 20 mL of standard buffer before elution with 9 mL of standard buffer with 25 mM *d*-desthiobiotin (Sigma). This elution fraction was bound to a 1 mL His-Trap™ column (GE Healthcare) equilibrated into standard buffer. The column was washed with 6% buffer B (standard buffer + 200 mM imidazole) for 5 CV to remove weakly bound impurities before elution of NTD-Strep with 10% buffer B. The contaminants were then eluted with 100% buffer B.

For a final polishing step, a Superdex 75 16/60 SEC column was equilibrated into standard buffer. Multiple 2 mL sequential applications were performed dependent on His-Trap™ elution volume with SEC flow reduced to 0.4 mL/min. Fractions containing NTD-Strep or dCTD were identified by SDS-PAGE, pooled and concentrated using a 10,000 Da MWCO concentrator (Vivaspin Vs2002) until protein concentration was above 10 mg/mL or volume was below 150 µL. Before aliquoting protein was centrifuged at 4°C, 14,000 RCF for 15 minutes and final protein concentration measured by absorbance at 280 nm using a Nanodrop™ spectrophotometer. Protein was flash-frozen in 10-20 µL aliquots and stored at -80°C for long-term storage.

2.2.4 Preparation of selenomethionine-labelled NTD_{31-145}-strep and NTD_{31-145}DM-Strep

The Molecular Dimensions minimal media base (MD12-501 / MD12-502) was used for production of selenomethionine-labelled protein. Initially 4 litres of Rosetta2™ DE3 cells expressing NTD_{31-145}-Strep or NTD_{31-145}DM-Strep were grown in 2XYT media to OD_{600} 0.6 using the standard procedure. Cells were harvested at 4,000 RCF for 10 minutes before suspending in 10 mL of autoclaved PBS, combined together and a further centrifugation at 4,000 RCF for 10 minutes. The cell pellet volume was adjusted to 45 mL with autoclaved PBS and distributed between the four 1 litre flasks of minimal media. Cells were grown for 40 minutes in minimal media at 25°C before addition of 50 mg of powdered selenomethionine (Arcos 259960025) per 1 L of minimal media. After a further 90 minutes protein production was initiated using 0.5 mM IPTG and temperature reduced to 20°C for 16 hours for protein production. Cells were subsequently harvested at 4,000 RCF for 10 minutes, prior to suspension in standard buffer supplemented with 3 mM TCEP and protease inhibitor tablets.

Protein was purified from 4 litres of cells using the standard procedure (2.2.3) with the following alterations. Initial capture was with a 5 mL Amylose column (NEB) pre-equilibrated into standard buffer + 3 mM TCEP, with 0.5 mM TCEP included throughout this procedure. Dialysis with TEV cleavage was performed using the standard protocol followed by Strep-Tactin capture with no reducing agent on elution. His-Trap™ chromatography was performed as standard with elution fractions applied in two sequential Superdex 75 16/60 SEC column runs at flow rate 0.4 mL/min pre-equilibrated into standard buffer either with or without 2mM DTT (sigma)

2.2.5 Purification of CTDex, MBP-CTD, CtIP-CTD and its mutants

For each purification, a resuspended pellet from 2 litres of cells was thawed and lysed by either sonication or high-pressure homogenisation (EmulsiFlex) before clarification at 45,000 RCF for 1 hour at 10°C, then filtered through a 5.0 µm membrane (Minisart 17594-K). For initial capture, a 5 mL Amylose (NEB) column was prepared and pre-equilibrated with the same resuspension buffer. Filtered supernatant was bound to the column by gravity flow dropwise, washed with 20 mL of standard buffer

before elution with standard buffer + 10 mM maltose. Protein was dialysed overnight into the standard buffer with addition of TEV were necessary to remove the His-MBP tag. TEV protease was pre-activated with 10 mM DTT (Sigma) for 30 minutes at 25°C before mixing with the eluted protein.

To remove the cleaved His-MBP tag from the protein preparation, the dialysed sample was first diluted to 100 mM NaCl with 20 mM Tris pH 8. Diluted sample was subsequently loaded onto a 5 mL Hi-Trap™ Heparin column (GE Healthcare) that had been pre-equilibrated with buffer A (20 mM Tris pH 8, 100 mM NaCl). The column was then washed with 10 mL of buffer A, before the flow was reversed and a further 10 mL of buffer A was washed over the column. Contaminants were removed with a step to 17.5% v/v buffer B (20 mM Tris pH 8, 1000 mM NaCl) for 15 mL before reducing the buffer B% to 0 over a further 10 ml. A gradient of buffer B from 0-60% was then used to elute the CTD protein and mutants. 100% v/v buffer B was used to wash the column of other bound proteins. For CTDex, a step gradient of 40% v/v buffer B was used to elute and concentrate the protein.

Elution fractions corresponding to CtIP-CTD were pooled and concentrated using a 10,000 MWCO PES concentrator (Vivaspin® vs2002) until the pooled volume was below 1 mL. Where indicated within the results a 200 µl amylose column was then prepared using a 1 mL gravity flow column (Expedeon AMCO25), pre-equilibrated into the standard buffer before allowing the concentrated CTD to pass through by gravity flow. The flow, along with a 200 µL wash fraction from the final amylose column (or the 1 mL concentrated fraction when this stage was omitted) was then spun at 14,000 RCF for 10 minutes 4°C before injection onto a pre-equilibrated Superdex 75 16/60 SEC column. The eluted fractions were collected. Fractions corresponding to CtIP-CTD were pooled and concentrated, with final concentration measured by absorbance at 280 nm. Protein was flash frozen in 10-20 µL aliquots and stored at -80°C for long-term storage.

2.2.6 Expression and purification of And1$_{1\text{-}826}$ and And1$_{316\text{-}826}$

Rosetta™ 2 (DE3) cells were transformed with an And1$_{1\text{-}826}$ or And1$_{316\text{-}826}$ construct using the standard procedure prior to growth in liquid media. Four litres of 2XYT media were used for each construct, grown until OD_{600} reached 0.8, and induced with 0.5 mM IPTG (Generon). Cells were incubated at 20°C before harvesting 16 hours post induction. Cells were resuspended using 20 mM HEPES pH 7.2, 500 mM NaCl, 10 mM imidazole and SigmaFAST™ EDTA-Free protease inhibitors before flash freezing in liquid nitrogen.

For each construct four litres of cells were lysed by high-pressure homogenisation (EmulsiFlex) (4 passes) before clarification at 45,000 RCF for 1 h at 10°C and filtration through a 5.0 µm membrane. A 2 mL Ni-NTA column was prepared using His-select® Ni-NTA affinity gel (Sigma Cat no. H0537) at 25°C

and equilibrated with the resuspension buffer. Protein was bound dropwise before washing with 12 CV of resuspension buffer and elution in 1 CV fractions with 20 mM HEPES pH 7.2, 500 mM NaCl, 200 mM imidazole. Fractions containing And1 were identified by SDS-PAGE and pooled for further purification by ion exchange chromatography using a 5 mL HiTrap™ Q column (GE healthcare). NaCl concentration was diluted to 160 mM for binding to the column, which had been pre-equilibrated in buffer A (20 mM HEPES pH 7.2, 100 mM NaCl). A linear gradient of buffer A to buffer B (20 mM HEPES pH 7.2, 1 M NaCl) was used to elute the protein. Fractions containing And1 were analysed by SDS-PAGE, pooled, and concentrated using a 10,000 Da MWCO PES concentrator (Vivaspin® VS2002). Purified $And1_{1\text{-}826}$ was flash frozen in liquid nitrogen for future pull-down experiments. Purified $And1_{316\text{-}826}$ was further polished by SEC using a Superdex 200 16/60 column. Resultant fractions corresponding to $And1_{316\text{-}826}$ were identified by SDS-PAGE, pooled and flash frozen in liquid nitrogen.

2.2.7 *E. coli* purification of $CtIP_{FL}$

A resuspended pellet from 2 litres of *E. coli* Rosetta™ 2 (DE3) cell expression of $CtIP_{FL}$ was thawed and sonicated using the standard procedure; cell debris were removed by centrifugation at 45,000 RCF for 1 hour at 10°C. Supernatant was filtered through a 5.0 µm membrane before loading onto a 2 mL Ni-NTA resin (Qiagen) bed dropwise by gravity flow. The column was washed with five 5 mL washes with standard buffer supplemented with 10 mM Imidazole, then eluted with five 3 mL washes of standard buffer + 200 mM Imidazole. $CtIP_{FL}$-containing fractions were identified by SDS-PAGE and Western Blot.

2.3 Bacmid generation for baculovirus insect cell expression

2.3.1 Transformation and selection

Dh10Bac cells (Gibco) were used for recombination of the pACEBac vector into the bacmid within the Dh10Bac cells for viral amplification. Cells were transformed using the standard protocol with 1 ng of an individual pACEBac1 protein expression construct (CtIP, Mre11, Rad50 and Nbs1) with 20 µL of DH10Bac cells. Following transformation cells recovered at 37°C for 5 hours with orbital shaking. Cells were then plated out onto LB-agar with selection (table 2.3.1). Plates were incubated for 48 h at 37°C until blue and white colonies could be distinguished. Individual white colonies were picked and plated on fresh LB-agar plates with the same selection for a further 24-48 hours at 37°C. Confirmed white colonies were then cultured for alkaline lysis.

Ingredients	**Stock solution/ mg per mL**	**Solvent**	**Final concentration/ µg per mL**	**Volume for 100ml of agar/ µL**
Kanamycin	34	Milli-Q water	50	150
Gentamycin	7	Milli-Q water	7	100
Tetracycline	5	Ethanol	10	200
X-Gal	20	DMSO	200	1000
IPTG	238.3	Milli-Q water	40	67

Table 2.3.1: **Selection of blue white colonies using Dh10Bac cells and pACEBac1.**

2.3.2 Alkaline lysis Ethanol precipitation

Cells were grown in the presence of the same antibiotics as in the LB-Agar (table 2.3.1) in 5 mL of 2xYT for 24-36 hours at 37°C with orbital shaking. 0.3 mL of the cells were removed for a glycerol stock and mixed with 0.3 mL of 50% v/v glycerol and kept at -80°C. The remaining cells were harvested at 5,000 RCF for 10 minutes before resuspension in 0.3 mL of buffer P1 (table 2.3.2), 0.3 mL of buffer P2 was added and the tube inverted 4-6 times to achieve lysis; this was followed by addition of 0.3 mL of buffer P3 and inversion 4-6 times for neutralisation. The mixture was then incubated on ice for 10 minutes before 10 minutes centrifugation at 14,000 RCF.

Buffer	**Composition**
P1	50 mM Tris-HCL pH 8.0 10 mM EDTA 100 µg/mL RNase A
P2	200 mM NaOH 1% w/v SDS
P3	3.0 M potassium acetate pH 5.5

Table 2.3.2: **Alkaline lysis buffer composition.**

2.3.3 Ethanol precipitation

Supernatant from alkaline lysis or DNA from large volume PCR reactions were precipitated by dropwise addition to 0.8 mL of isopropanol (Sigma) and incubation on ice for 10 minutes. DNA was pelleted by centrifugation at 14,000 RCF for 15 minutes, washed with 0.5 mL of cold 70% v/v ethanol (Sigma) and re-pelleted at 14,000 RCF for a further 10 minutes before the supernatant was removed. The pellet was air-dried for 20 minutes at room temperature. It was then dissolved in TE buffer.

2.3.4 Bacmid PCR confirmation and Colony PCR

For colony PCR the starting material was either purified bacmid (1.5 µL) or a single colony from a transformation plate. For a single colony the cells were mixed into 10 µL of LB-broth which contained 25% v/v glycerol. 1 µL of this was then used for the 13 µL PCR reaction using the M13 primers. Cycle conditions were 98°C for 5 minutes followed by 35 cycles of 98°C for 30 seconds, 55°C for 30 seconds, 72°C for 30 seconds per 1 kb of DNA with final extension for 10 minutes at 72°C. Products were analysed by 1% w/v agarose gel electrophoresis.

2.4 Recombinant protein production in *Sf9* insect cells

Sf9 insect cells were grown in Lonza insect express medium (Cat NO. BE12-730Q) at 27°C with constant 120 RPM orbital shaking, 25 mm throw, (no humidity or CO_2 control) continually with passage number recorded. Stock cells were maintained in flat-bottomed flasks (Table 2.4). Protein expression was performed in 1 litre of cells per 2 litre baffled-bottom flask. Cell density was maintained between $0.5x10^6$ and $5x10^6$ cells per mL with cell viability monitored. The passage number of the cells was recorded, and the cells were split on Mondays, Wednesdays and Fridays until the passage number was above 25 or the cell viability reduced to below 80%. Frozen cell stocks were maintained to allow constant availability of cells for expression.

Flask size	**Culture volume/ mL**
125 mL – Corning 431143, Triforest 355117B	20-50
250 mL – Corning 431144	50-100
2L Baffled – Corning 432156, triforest 355131B	1000
2L Non-baffled – Corning 431255, triforest 355133B	Insect cells 200-800 HEK293F cells 300-400ml

Table 2.4: **Flask details for HEK293F and insect cell expression.**

All protein purification stages were carried out at 4°C unless stated otherwise, with buffers filtered and degassed using a 0.22 µm membrane (Triple Red FPE214500).

2.4.1 Expression and purification of the MRN complex

Plasmids for baculovirus expression in insect cells of Flag-tagged Mre11, His_6-HA tagged Rad50 and Nbs1 were donated by Tanya Paull (Lee and Paull, 2006). All purification stages were carried out at 4°C with the MRN buffer containing 50 mM phosphate buffer pH 7.0, 100 mM KCl, 10% v/v glycerol filtered and degassed using a 0.22 µm membrane (triple red FPE214500).

One litre of *Sf*9 cells were transfected at a cell density of $2x10^6$ cells per mL with a Multiplicity of infection (MOI) ratio of 2.0 Mre11, 6.0 Rad50 and 4.0 Nbs1 bacmids. Cells were harvested by centrifugation at 500 RCF for 10 minutes at 3 days post infection, washed in PBS and the cell pellet flash frozen in liquid nitrogen. The cell pellet was thawed and resuspended in 70 mL of MRN buffer with additional 0.5% v/v Tween 20 and protease inhibitor tablets (SigmaFAST™ EDTA free). Cells were lysed by high-pressure homogenisation (EmulsiFlex) (4 passes) and lysate clarified at 75,000 RCF for 1.5 hours, 10°C using polypropylene centrifuge tubes (Nalgene CAT NO: SQ-3119-0050). Clarified lysate was filtered sequentially through a 5.0 µm membrane syringe filter (Minstart cat no 514-4129) and a 0.45 µm membrane vacuum filter (Milipore cat no. SCHVU01RE) to remove any large particles and to de-gas the sample. A 20 mL 16/10 Fast flow Hi-Prep™ heparin column (GE Healthcare) was pre-equilibrated with MRN lysis buffer. Filtered lysate was bound to the heparin column by sample pump

at 2.0 mL/min prior to washing with MRN buffer for 40 mL in forward flow, and 40 mL in reverse. Protein was eluted in reverse flow with 50% MRN buffer B (MRN buffer with final concentration 1 M KCl) in a 1 CV peak. For the next step of Flag-affinity purification, a 1 mL M2 anti-Flag affinity column (Sigma Cat no. A2220) was prepared in MRN buffer. The combined heparin elution salt concentration was diluted to 100-200 mM and washed twice over the anti-Flag column with gravity flow dropwise. The column was washed four times with 10 mL of MRN buffer. Protein was eluted by overnight incubation with 2 mL of MRN buffer containing 1 mg/mL Flag peptide at 4°C with constant rolling. A second elution fraction was obtained by washing further with 2 and then 1 mL of elution buffer. The eluted samples were combined and bound to a 5 mL Hi-Trap™ S column (GE Healthcare) equilibrated into MRN buffer for removal of Flag peptide, and the protein was eluted with MRN buffer containing 500 mM KCl. Elution fractions from Hi-Trap™ S were combined and buffer exchanged into MRN buffer using a PD10 column (GE Healthcare). Protein was concentrated to 200 µL using a 30,000 MWCO concentrator (Vivaspin) with concentration determined by Cary UV absorbance at from 220-360 nm using a theoretical extinction coefficient of 159,980 M^{-1}. A concentration of 0.15 mg/mL in 200 µL of sample volume was normally achieved per 1 L of insect cell culture.

2.4.2 Expression and purification of And1

*Sf*9 cells at 2.0x10^6 cells per mL were transfected with 3 mL of recombinant baculovirus per litre of cells. Baculovirus was generated by Dr Mairi Kilkenny. Cells were harvested 3 days post transfection by centrifugation at 500 RCF for 10 minutes, followed by washing with 1X PBS and a further 500 RCF centrifugation. Cells were frozen in liquid nitrogen and stored at -80°C for storage.

One litre of *Sf*9 cells were thawed and suspended in 100 mL of And1 lysis buffer (25 mM HEPES pH 7.2, 500 mM KCl, 5% v/v glycerol, SigmaFAST™ EDTA-Free protease inhibitors, Benzonase® before lysis by high-pressure homogenisation (EmulsiFlex) (4 passe). Cell lysate was clarified by centrifugation at 45,000 RCF for 1 h at 10°C followed by vacuum filtration through a 0.45 µm membrane (Millipore cat no. SCHVU01RE). Clarified lysate was bound to a pre-equilibrated 3 mL Ni-NTA column (Sigma H0537) and washed with And1 wash buffer (25 mM HEPES pH 7.2, 500 mM KCl, 5% v/v glycerol, 25 mM imidazole). Elution was performed in 1 CV fractions using And1 elution buffer (25mM HEPES pH 7.2, 500 mM KCl, 5% v/v glycerol, 200 mM imidazole). Fractions containing And1 were confirmed by SDS-PAGE and subsequently pooled, then supplemented with 1 mM EDTA and incubated overnight at 4°C. Protein salt concentration was diluted to 200 mM KCl and applied to a 5 mL Hi-Trap heparin™ column (GE Healthcare) and eluted with a linear gradient from 100 mM to 1M KCl. Fractions containing And1 were confirmed by SDS-PAGE, pooled and concentrated to 1 mL using a 30,000 MWCO concentrator (Vivaspin® VS2022). Concentrated protein was applied to a 16/60 Superdex 200 SEC column, pre-

equilibrated into 25 mM HEPES pH 7.2, 140 mM KCl, 0.5 mM TCEP. Fractions containing And1 were detected by SDS-PAGE, pooled and concentrated to 1.2 mg/mL. Protein aliquots were frozen in liquid nitrogen before storage at -80°C.

2.4.3 Baculovirus expression of $CtIP_{FL}$ and L27E $CtIP_{FL}$

$CtIP_{FL}$ P2 virus was filtered and frozen in 1 mL aliquots for amplification of P3 virus for protein expression. Fresh P3 virus was added at a MOI of 2.0 into 50 mL to 2 L of *Sf9* cells at $2.0X10^6$ cells per mL with constant 120 RPM shaking, 25 mm throw at 27°C with a viability of ≥95% that was monitored daily. Cell harvesting occurred at day 2 by centrifugation at 500 RCF for 10 minutes, with supernatant removed and cell pellet washed with PBS prior to freezing.

A cell pellet from 0.5 to 2.0 litres of *Sf*9 cells that had expressed $CtIP_{FL}$ was thawed and suspended in 35-140 mL final volume of lysis buffer (20 mM Tris pH 8, 100 mM NaCl, 0.5 mM TCEP pH 8, 10% v/v glycerol, 0.5% v/v Tween 20, SigmaFAST™ EDTA-Free protease inhibitors, 2.5 units of Benzonase® (Sigma)). Cells were lysed by high-pressure homogenisation (EmulsiFlex) (4 passes) and lysate clarified at 75,000 RCF for 1.5 hours, 10°C using polypropylene centrifuge tubes (Nalgene CAT NO: SQ-3119-0050). Clarified lysate was filtered sequentially through a 5.0 µm membrane syringe filter (Minstart cat no 514-4129) and a 0.45 µm membrane vacuum filter (Milipore cat no. SCHVU01RE) to remove any large particles and to de-gas the sample. Filtered lysate was bound to a 1 mL M2 anti-Flag affinity column (Sigma Cat no. A2220) by gravity flow dropwise and washed in *Sf9*-CtIP buffer (20 mM Tris pH 8, 100 mM NaCl, 0.5 mM TCEP pH 8 and 10% v/v glycerol). Protein was eluted by overnight incubation with 2 mL of *Sf9*-CtIP buffer containing 1 mg/mL Flag peptide at 4°C with constant rolling. A second elution fraction was obtained by washing further with 2 and then 1 mL of elution buffer. Fractions containing CtIP were confirmed by SDS-PAGE, pooled and concentrated to 0.5 mL using a 30,000 MWCO concentrator (Vivaspin® VS2022). Concentrated protein was applied to a Superdex 200 10/300 SEC column, pre-equilibrated into *Sf9*-CtIP buffer. Fractions containing $CtIP_{FL}$ were detected by SDS-PAGE. Protein was concentrated to 400 µL using a 30,000 MWCO concentrator (Vivaspin® VS2022) with concentration determined to be 0.7 mg/mL or 6.75 µM by Cary UV absorbance at from 220-360 nm using a theoretical extinction coefficient of 41715 M^{-1}. Protein aliquots were frozen in liquid nitrogen before storage at -80°C.

2.5 Recombinant protein production of $CtIP_{FL}$ from HEK293F cells

Human Embryonic Kidney 293 freestyle (HEK293F) cells (Invitrogen) were grown in Freestyle™ media (Invitrogen) in a Kuhner incubator at 37°C, 8% CO_2, 100% Humidity, 125 RPM shaking, and 25 mm throw, continually with passage number recorded. Flat bottomed flasks detailed in table 2.4 were used. Cells were split Mondays and Fridays to $0.2x10^6$ cells per mL in 25 mL to maintain cells in the growth phase until the passage number reached above p15 or cells began to form clumps. At this point a fresh batch of cells would be started from a frozen cell stock.

2.5.1 DNA purification for HEK expression

For high quantities of DNA for transfection a QIAGEN® Mega prep (endotoxin free – Cat No. 12381) was used, yielding around 1 to 3 mg of plasmid DNA from 1 Litre of cells. The sequence confirmed pACEMam1 $CtIP_{FL}$ DH10β glycerol stock was inoculated into 50 mL of 2xYT with gentamicin. The cells were grown at 37°C for 8 hours before inoculation into 1 Litre of LB media with gentamicin. The 1 litre culture was grown at 30°C 180 RPM overnight before harvesting by centrifugation at 4000 RCF for 10 minutes. The manufacture's protocol was then followed for the remainder of the purification with a few alterations. Firstly, on binding to the column (step 13) the remainder of the filtered DNA was kept on ice. In step 16 the eluted DNA was chilled prior to precipitation, with the centrifugation (step 17) increased to 30 minutes to ensure the DNA was precipitated.

2.5.2 HEK293F protein expression

For protein expression, cells were split until the required volume was achieved at $0.5x10^6$ cells per mL on the day prior to transfection. The following day cells were counted using a nebular chamber, and the total number of cells was calculated. To transfect the cells, Freestyle™ media (Invitrogen) corresponding to 1/10th of the total culture volume was used; 1.2 µg of pACEMam Twin-Strep $CtIP_{FL}$ DNA along with 2.4 µg PEI were added per 1 million cells. The mixture was vortexed for 10 seconds and incubated at room temperature within the hood for 15 minutes before addition to the cells that were to be transfected. Three to four hours post transfection a final concentration of 4 mM filtered valproic acid was added to the transfected cells. The length of expression was experimentally determined by removal of 1 mL of cells daily with total and soluble protein analysed by Western Blot. Valproic acid-treated cells were harvested four days post transfection at 500 RCF for 10 minutes, washed in 1X PBS, and harvested at 500 RCF for 10 minutes with supernatant removed. Cell pellets were flash frozen in liquid nitrogen prior to storage at -80°C.

All protein purification stages were carried out at 4°C unless stated otherwise, with buffers filtered and degassed using a 0.22 µm membrane (Triple Red FPE214500).

2.5.3 $CtIP_{FL}$ HEK293F Protein purification

A cell pellet from 0.4-1.2 L of HEK293F cells that had expressed $CtIP_{FL}$ was thawed on ice for 1 hour prior to resuspension in 20 to 40 mL final volume of lysis buffer (50 mM Tris pH 8, 100 mM NaCl, 0.5 mM TCEP pH 8, 10% v/v glycerol, 0.5% v/v Tween 20, SigmaFAST™ EDTA-Free protease inhibitors, 2.5 units of Benzonase® (Sigma)). Lysis by sonication, pulsing with 5 seconds on, 10 seconds off for 3 minutes with 60% amplitude was performed at 0.5°C using an ice water bath. To pellet cell debris the lysate was equally distributed into two polypropylene centrifuge tubes (Nalgene™ CAT NO: SQ-3119-0050) and centrifuged at 70,000 RCF for one and a half hours at 10°C (to limit the temperature increase the rotor was precooled to 5°C at 50,000 RCF). Following centrifugation, the supernatant was filtered sequentially through a 5.0 µm membrane syringe filter (Minisart cat no 514-4129) and a 0.45 µm membrane vacuum filter (Milipore cat no. SCHVU01RE) to remove any large particles and to de-gas the sample. Filtered lysate was loaded at 0.1 mL/min via 50 mL Superloop onto a pre-equilibrated with lysis buffer 1 mL StrepTrap™ HP (GE healthcare) column controlled by an AKTA purifier at 10°C. The column was further washed with lysis buffer until UV reached the baseline level, followed by reversing the flow and washing with HEK293F-CtIP buffer (50 mM Tris pH 8, 200 mM NaCl, 0.5 mM TCEP) for 10 CV. Protein was eluted in a sharp peak using standard buffer with additional 25 mM *d*-desthiobiotin (Sigma), with pH adjusted to 8. Fractions containing $CtIP_{FL}$ were identified by SDS-PAGE before further purification, preparation of Cryo EM grids, or flash freezing for experimental analysis. Protein concentration was determined by UV absorbance at 280 nm using concentrator flow through as a blank due to buffer background absorbance.

Buffer exchange was achieved by two methods: SEC and NAP-5 (GE Healthcare) desalt columns. For SEC either GE Superose 6 3.2/300 increase or GE Superdex 200 10/300 column was used, and 50 or 200 µL fractions were collected. For NAP-5 buffer exchange the manufactures protocol was used, collecting 200 µL fractions on elution. Both methods buffer exchanged $CtIP_{FL}$ into HEK293F-CtIP buffer for analysis by SDS-PAGE and protein concentration estimation by absorbance at 220-360 nm.

2.6 SDS-Page Protein Analysis

10 µL of protein solution was mixed with 5 µL of loading buffer (27 µL of 10X BOLT® reducing agent (Invitrogen) with 63 µL of 4X NuPage® LDS buffer (Invitrogen)). Following heat treatment at 100°C for 1-minute 10 µL of the resulting sample was loaded onto a 12% Bis-Tris gel or 4-12% gradient gel (Invitrogen), together with either See-Blue® or See-Blue® Plus 2 pre-stained protein marker in the first lane (Invitrogen).

To prepare 12% Bis-Tris gels, 1 mm BOLT® gel cassettes (Invitrogen) were filled with 12% separating gel (diluted from 30% v/v acrylamide solution), 0.36 M Bis-Tris pH 6.8, 0.11% TEMED (v/v) and 0.035%

ammonium persulphate (w/v), overlaid with isopropanol and allowed to set. After 1 hour, isopropanol was poured off and allowed to evaporate before addition of 4% stacking gel (diluted from 40% acrylamide), 0.36 M Bis-Tris pH 6.8, 0.35% TEMED (v/v) and 0.1% ammonium persulphate (APS) (w/v) with either a 12 or 15 lane comb inserted. Finished gels were stored at 4°C with 1X MOPS running buffer.

Polyacrylamide gels were run in BOLT® 1X MOPS running buffer (Invitrogen) at 200 V for 33 minutes. Before staining, gels were washed in 50 mL of distilled water with heating for 30 seconds. This was followed by staining with SimplyBlue™ SafeStain (Invitrogen): 50 mL of stain was added to the gel and microwaved for 40 seconds on high before gentle rocking for approximately 5 minutes. This was repeated 2-3 times until bands were visible. De-staining was performed with distilled water.

For western blot analysis the Bolt mini blot module was used using manufactures guidelines and nitrocellulose membranes (0.45 µm pore size). Antibodies were used according to manufactures protocol, see appendix 1 supplementary figure 7.6 for antibodies used.

2.7 Nuclease assay

The nuclease assay was performed as described by Sartori *et al.* 2007. The PhiX174 circular single-stranded virion DNA substrate (5,386 nucleotides, New England Biolabs) was mixed with 40 ng of purified Mre11-Rad50-Nbs1 together with 200 ng BSA. For HEK293F $CtIP_{FL}$, either 200 or 350 ng of CtIP was used per reaction. Reaction buffer contained 30 mM potassium-MOPS, pH 7.2, 1 mM dithiothreitol, 1 mM ATP, 25 mM KCl and 5 mM of either $MgCl_2$ or $MnCl_2$. After incubation at 37 °C for 3 h, nuclease reactions were terminated by adding 0.1 volume of stop solution (3% SDS w/v, 50 mM EDTA) and samples were run in a 0.8% w/v agarose gel (1× TAE) for 90 min at 100 mA, 4°C. DNA species were stained with SYBR Gold Nucleic Acid Gel Stain (Molecular Probes) for 20 min and visualized using the UV trans-illuminator (UVP).

2.8 Thermal melt analysis

SYPRO® Orange binds non-specifically to hydrophobic surfaces, with water strongly quenching its fluorescence (Huynh and Partch, 2015). When a protein unfolds, any hydrophobic surfaces that are now exposed bind the dye, resulting in an increase in fluorescence by excluding water. This technique can be used with increasing temperature to denature the protein and a measurement of the resultant fluorescence plotted to calculate the melt peak.

For thermal melt analysis Melt curves were generated using the Bio-Rad CFX Connect real time PCR machine with temperature increased by 1.0°C every 30 seconds and fluorescence recorded. Two concentrations of $CtIP_{FL}$ were used: 8 µM and 5 µM, and dilutions were made with the standard buffer.

Final dilution of SYPRO® Orange was 1:2000. The NTD protein concentration was optimised to 12.5 µM for the buffer screen (Appendix figure 7.1), with 12.5 µL of protein mixed with 12.5 µL of buffer to be assessed with final concentration 1:2000 SYPRO® Orange. For thermal melt analysis of the NTD cluster mutations concentrations of 16, 32 and 64 µM of protein were used in 25 µL with final concentration 1:2000 SYPRO® Orange.

2.9 Spectrophotometric analysis of zinc content

To assess the concentration of zinc in a protein sample, a 30 µM 4-(2-pyridylazo)-resorcinol (PAR) and a zinc acetate standard curve were generated. For the standard curve, known concentrations of zinc acetate were diluted from a 1 M zinc acetate stock (Sigma) using filtered MiliQ water. Sequential 10x dilutions were made from 1 M to 1 mM zinc acetate, with this used to make a 50 µM stock. The 50 µM stock used to generate a series of zinc standards from 5-50 µM in 5 µM steps. For the PAR, a 140 mM stock was prepared in DMSO with 3 µL diluted in 14 mL of protein buffer for 30 µM working stock. For the baseline reading, 3 µL of DMSO was added to 14 mL of protein buffer. 10 µL of the 5-50 µM zinc acetate solutions were incubated with 80 µL of 30 µM PAR for 5 minutes before the UV spectra were measured over a range of 300-600 nm (Varian Cary 50 spectrophotometer). The absorbance of each peak at 492 nm and 414 nm was recorded. Plotting of A492/(A492+A414) against zinc acetate molar concentration was generated as a standard curve. (Pfister et al., 2000).

The purified proteins (in 20 mM Tris pH 8, 300 mM NaCl buffer) were diluted to 0.5 to 1 mg/mL depending on the expected yield of zinc, with accurate protein concentration determined by UV absorbance at 280 nm. 5 µL of protein was removed for SDS page before protein samples were digested in a total volume of 80 µL with 50 µg of protease K (Sigma) at RT overnight with gentle agitation (or 60°C for 1 hour) followed by centrifugation at 14,000 RCF for 10 minutes at 4°C. The supernatants were collected for zinc analysis. A further 5 µL of supernatant was removed for a post-digestion SDS-PAGE sample.

For each sample, 10 µL of supernatant was mixed with 80 µL of 30 µM PAR solution and incubated at room temperature for 5 minutes. A spectrum was taken and the value of A492/(A492+A414) was calculated compared with the standard curve. The determined zinc molar concentration and known protein molar concentration together gave a ratio of protein to zinc. Graphs were generated in Microsoft Excel.

2.10 SEC-SAXS

Size Exclusion Chromatography Small Angle X-Ray Scattering (SEC-SAXS) is a powerful technique that can be used to obtain structural information about ordered and disordered regions of biological samples to a low resolution (Feigin and Svergun, 1989). It can provide information about the shape and size of proteins as well as structural changes that can occur from mutation and addition of protein tags. The high-throughput of samples under physiological conditions can provide structural characterisation to compliment other techniques (Rambo and Tainer, 2013). X-rays are scattered by electrons that are part of an individual macromolecule. When a sample is homogeneous this collection of scattered electrons defines the molecular form factor A(q). The molecular form factor determines the amplitude of the scattered X-rays at a scattering vector, q, by an object with fixed orientation and electron density, p(r):

$$A(q) = \int p(r)\ x\ e^{iqr}\ dV$$

The scattering integral is evaluated at a distance **r** throughout the particle volume **V**, therefore an observed scattering intensity **I(q)** depends directly on the paired distance between all electrons in the particle. This gives rise to the electron paired-distance distribution function **P(r)** (Rambo and Tainer, 2013). From this **P(r)** distribution the maximal distance between two electrons within the molecule (D_{MAX}) can be calculated.

SEC-SAXS experiments were performed at Diamond Light Source on beamline B21 under the supervision of Dr Robert Rambo and Dr Nathan Cowieson. Default settings were used for the B21 beamline set up. Camera length is fixed at 4.014 meters with an energy of 12.4 Kev and flux of $x10^{12}$ photons per second.

Protein samples were taken to DLS on dry ice. After thawing, tubes were centrifuged briefly, combined and diluted with standard buffer were necessary before centrifugation at 14,000 RCF for 10 minutes. For all proteins the standard 20 mM Tris pH 8 300 mM NaCl buffer was used. A minimum of 50 µl of each sample was dispensed into a Greiner V-bottom 96 well plate (651201) sealed with an X-pierce™ cross slit seal (Sigma Z722529) and stored within the HPLC system at 4°C prior to injection. SEC was performed by the Agilent 1200 HPLC system with a Superdex 200 increase 3.2/300 2.4 mL column (GE Healthcare) at a flow rate of 0.05 or 0.07 mL/min with automatic sample loading. 35 µL of sample was used in session 1, and 45 µL of sample was used in session 2 for each SEC run for SEC-SAXS experiments. Details of proteins used are supplied in tables 2.10-1 and 2.10-2.

Protein	Concentration mg/ mL	Flow rate / mL per min
NTD_{30-145}	15.70	0.05
NTD_{31-145}	20.14	0.05
NTD_{30-145}Strep	4.62	0.05
NTD_{31-145}Strep	12.47	0.05
MBP_{30-145}strep	12.50	0.05
NTD_{31-145}C1-Strep	12.47	0.05
NTD_{31-145}C2-Strep	13.13	0.05
NTD_{31-145}C3-Strep	10.63	0.05
MBP	6.97	0.05

Table 2.10-1 **Protein concentrations for NTD SEC-SAXS.** Session 1.

Protein	Concentration mg/ mL	Flow rate / mL per min
NTD_{31-145}DM-strep	9.80	0.05
NTD_{31-145}Strep	9.93	0.05
NTD_{18-145}	4.42	0.05
NTD_{31-88}Strep	3.69	0.05
NTD_{93-145}Strep	5.86	0.05
MBP	6.97	0.05 and 0.07 separately
CTDex – B	2.86	0.07
dCTD	3.86	0.07

Table 2.10-2 **Protein concentrations for NTD SEC-SAXS.** Session 2.

2.10.1 SCÅTTER analysis of SAXS data and DAMIN/DAMIF model generation

The X-ray scattering data were processed using the B21 automatic processing pipeline. SCÅTTER (version 3.0) was used to determine the radius of gyration (R_G) and D_{MAX} values based on Guinier analysis and the P(r) distribution function as outlined in the Bioisis.net tutorial. The first horizontal portion of the SEC trace was used for baseline subtraction. Models that fit the SAXS data were generated from the P(r) distribution and D_{MAX} values using 23 rounds of DAMMIF followed by averaging by DAMAVA to generate a DAMFIT file for modelling, these three programs form part of the ATSAS (Petoukhov et al., 2012) package and also covered in the bioisis.net tutorial. SAXS figures were generated in PyMol and overlaid manually.

2.11 Mass spectrometry analysis

Purified protein samples in solution and protein species separated by SDS-PAGE where indicated were analysed by MALDI-TOF MS (Matrix-Assisted Laser Desorption/Ionisation – Time of Flight Mass Spectrometry) by Dr Len Packman (Protein & Nucleic Acid Chemistry Facility (PNAC), University of Cambridge). Amino Acid Analysis (AAA) was performed on purified homogeneous proteins at full and half concentrations by Peter Sharratt (PNAC Facility, University of Cambridge).

2.12 Semi-analytical analysis of CtIP-CTD

MBP-fused CtIP-CTD protein was purified using the standard procedure in 2.2.5 until after the heparin chromatography stage with removal of the TEV cleavage. From here, protein containing MBP-CTD was concentrated to 1 mL using a 30,000 MWCO PES concentrator (Vivaspin® VS0221). SEC was performed at room temperature using a Superdex 200 10/300 SEC column (GE), pre-equilibrated with standard buffer. For each experiment, a 100 µL of sample (90 µL of protein and 10 µL of standard buffer with 10X additive) was applied to the column. Traces were overlaid using the Unicorn software.

2.13 Cryo-EM

Grids (Quantifoil R 1.2/1.3 Cu 300 mesh) were glow discharged using a Pelco Easiflow device using manufacturer's instructions under the following settings: 25 mA, 0.39 mBar, 2 minutes, and carbon side up. Following glow discharge on a single side, 3 µL of freshly purified $CtIP_{FL}$ (Nap5 buffer exchanged protein into 50mM Tris pH 8.0, 300 mM NaCl, 0.5 mM TCEP at 0.9 mg/ml (monomer)) was applied and snap frozen in liquid ethane using a Vitrobot Mark IV (FEI) device using manufacturer's instructions under the following settings: grids were blotted for 2.5 seconds at 4°C and 100% humidity, with a 'blot force' of -5 and no pause between sample application and blotting. The grids were visualised in the Titan Krios G2 microscope of the Cryo-EM facility at the department of NanoScience University of Cambridge. The microscope was prepared by Dr Dima Chirgadze with the Falcon 3 detector on integration mode and controlled by Dr Jamie Blaza.

For MBP-NTD_{18-145} the same grids, settings and procedures were used however with a 30 mA glow discharge. Protein was applied to the grid in buffer containing 150 mM NaCl, 20 mM Tris pH 8.0 at a protein concentration of 0.75 mg/mL (monomer). The grids were visualised in the Talos Arctica™ microscope of the Cryo-EM facility at the Department of Biochemistry, University of Cambridge. The microscope was prepared by Dr Dima Chirgadze with the Falcon 3EC detector and controlled by Dr Neil Rzechorzek.

2.14 Analytical Ultra Centrifugation (AUC)

Under ultra-centrifugal conditions the sedimentation of a protein occurs and is governed by three factors: (1) the centrifugal force - this is set depending on the predicted molecular mass of the protein and determined by the speed of rotation; (2) the buoyancy force, which is proportional to the mass of the displaced solvent, as well as the speed of rotation; (3) the frictional force, which is governed by the frictional coefficient and the sedimentation velocity of the protein. The balance of these three factors at a constant ultra-centrifugal velocity allows determination of the sedimentation coefficient. This coefficient remains constant incorporating both the shape and molecular mass of the measured protein (Mächtle and Börger, 2006; Scott et al., 2005). In general proteins with an asymmetrical structure and smaller proteins sediment slower than those of higher mass and compact globular structure which are fast to sediment.

A fresh $CtIP_{FL}$ purification was performed from HEK293F cells. Fractions containing the bulk of $CtIP_{FL}$ were pooled and dialysed against 2 litres of the standard gel filtration buffer overnight before centrifugation at 14,000 RCF, 4°C for 10 minutes. Protein concentration was measured using the Nanodrop™ spectrophotometer and was calculated to be 1.3 mg/mL; two dilutions were made at 50% and 25% concentration with the dialysis buffer.

The experiment was performed using an analytical centrifuge (Beckman Optima XL-A/1) at 40,000 RPM pre-equilibrated to 20°C. 400 µL of the dialysis buffer was dispensed into chamber 1 and an equal volume of protein sample dispensed into chamber 2 for analysis by sedimentation velocity analytical ultracentrifugation. Both interference and absorbance data were collected with assistance from Dr Katherine Stott. Protein partial specific volume, buffer density and buffer viscosity were calculated using the program Sednterp (Hayes, Laue, & Philo, 1995). Sedimentation distribution was analysed by fitting data to the Lamm equation using Sedfit (Brown & Schuck, 2008; Eisenberg, 2003; Schuck, 2000).

2.15 DNA oligo design and annealing

Substrates for DNA binding were designed using the online server NUPACK for generation of specific secondary structure and to ensure the double stranded DNA ends were stable at room temperature. Oligoes were ordered from IDT and are shown in table 2.15. DNA oligoes were annealed by heating to 9°C for 3 minutes before cooling by 0.5°C every 30 seconds until reaching 4°C. Final concentration of all annealed substrates was 50 µM; the substrates were stored in black opaque tubes at -20°C.

Oligo	**Sequence**
FdX60	[Flc]ATGGTGTGTGTAGGTTAATGTGAGGAGGAGAGGTGAAGAAGGAGGAGAGAAGAAGGAGGC
CdX60	GCCTCCTTCTTCTCTCCTCCTTCTTCACCTCTCCTCCTCACATTAACCTACACACACCAT
20bp_NuPac_FWD	[Flc]GTGTGTGGTGGTGTTAGTTG
20bp_NuPac_REV	CAACTAACACCACCACACAC
40bp_NuPac_FWD	[Flc]ACCCCAACCTAAAATCCCACACCAACACACAACCCCCACT
40bp_NuPac_REV	AGTGGGGGTTGTGTGTTGGTGTGGGATTTTAGGTTGGGGT
FdX35	[Flc]ATGGTGTGTGTAGGTTAATGTGAGGAGGAGAGGTG
CdX35	CACCTCTCCTCCTCACATTAACCTACACACACCAT
FdX30	[Flc]ATGGTGTGTGTAGGTTAATGTGAGGAGGAG
CdX30	CTCCTCCTCACATTAACCTACACACACCAT
FdX25	[Flc]ATGGTGTGTGTAGGTTAATGTGAGG
CdX25	CCTCACATTAACCTACACACACCAT
FdX10	[Flc]GCCTCCTTCT
CdX10	AGAAGGAGGC
36bp_FdX60	GCCTCCTTCTTCTCTCCTCCTTCTTCACCTCTCCTCCTCACATT
16bp_FdX60	GCCTCCTTCTTCTCTC
44bp_FdX60	CTCCTTCTTCACCTCTCCTCCTCACATTAACCTACACACACCAT
14bp_FdX60	GCCTCCTTCTTCTC
24bp_FdX60	CTCACATTAACCTACACACACCAT
22bp_FdX60	CACATTAACCTACACACACCAT

Table 2.15: **Fluorescent oligoes.**

For the 200 bp substrate, primers [Flc] CCGAAGGTAACTGGCTTCAG and GCCTTATCCGGTAACTATCGTC were used to PCR amplify a dsDNA region spanning 200 bp from Puc19. Eight 50 µL, 40 cycle, PCR reactions were used per purification with an annealing temperature of 64°C and minimal Puc19 template (<10 pg). DNA was purified by ethanol precipitation (section 2.2.3).

2.16 Electrophoretic Mobility Shift Assay

Buffer concentrations were constant for each Electrophoretic Mobility Shift Assay (EMSA). For EMSAs in Chapter 5 CtIP full-length the buffer contained 20 mM Tris pH 8, 300 mM NaCl, 0.5 mM TCEP and 10% v/v glycerol. For EMSAs in Chapter 4 CtIP-CTD the buffer contained 20 mM Tris pH 8, 150 mM NaCl, either with or without 50 µM zinc acetate. Serial dilutions of CtIP-CTD, $CtIP_{FL}$ and L27E $CtIP_{FL}$ were generated maintaining buffer concentrations. 9 µL of protein was mixed with 1 µL of

fluorescently labelled DNA to give a final DNA concentration of either 1 µM or 100 nM within the EMSA. Protein:DNA mixture was incubated for a minimum of 20 minutes at 4°C. Separation of protein-DNA complexes from free DNA was performed by electrophoresis using a 1x Tris Borate (TB) ultrapure agarose gel for 3 hours at 25V, 4°C. DNA was visualised by UV trans-illumination or scanned using a Typhoon FLA 9000 imager (GE healthcare) at 473 nm.

2.17 Fluorescence polarisation

The theory of FP comes from observations that when small fluorescent molecules (tracers) are excited with plane-polarized light, the emitted light is largely depolarized because the molecules tumble rapidly in solution during their fluorescence lifetime (the time between excitation and emission). However, if bound by a larger molecule, the tracer's effective molecular volume is increased. The tracer's rotation is therefore slowed with more of the emitted light in the same plane as the excitation energy. The bound state therefore has a high intrinsic polarization value verses a low value for the free state. The measured polarization is a weighted average of the two values, thus providing a direct measure of the fraction of tracer bound to receptor. In this assay, the DNA was labelled with fluorescein and the anisotropy of this fluorescent species determined. As protein is added, it can bind the labelled species to produce a complex that tumbles more slowly and increases the anisotropy of the fluorophore. Analysis of the anisotropy change as a function of protein concentration produces a binding isotherm from which shows the formation of the protein-DNA complex (Lundblad et al., 2008; Rossi and Taylor, 2011).

The CtIP-CTD protein buffer was exchanged into 20 mM Tris pH 8, 150 mM NaCl with or without zinc acetate as indicated, followed by a 2-hour incubation on ice and centrifugation at 14,000 RCF for 10 minutes at 4°C. The CtIP-CTD concentration was quantified by 220-360 nm absorbance measured by a Cary-UV spectrophotometer using a theoretical extinction coefficient of 14690 M^{-1}. A serial dilution was used to generate the protein concentrations, retaining a consistent zinc concentration in the buffer. Protein samples were incubated with annealed 10 nM linear 5' Fluorescein-dsDNA 60 bp substrate, in 20 mM Tris pH 8, 150 mM NaCl with or without zinc acetate. Protein-DNA mixtures were transferred to black 96-well half-area plates (Corning Cat No. 3686). Fluorescence anisotropy data were recorded at 25°C using a PHERAstar FS plate reader (BMG Labtech), with fluorescence polarisation optics module exciting at 485 nm and recording emissions at 520 nm. Data were plotted using Microsoft Excel.

2.18 SEC-MALS analysis of CtIP-CTD

The theory behind SEC Multi Angle (Laser) Light Scatter (SEC-MALS) is demonstrated by the Debye-Zimm equation (Eq.2.18-1). The equation describes the relationship between the angular dependence of light scattered from a solution containing homogeneous protein and the molecular mass of that protein.

Eq.2.18-1

$$\frac{KC}{R_\theta} = \frac{1}{MP_{(\theta)}} + 2BC + 3BC^2 + \ldots$$

In this equation $R_{(\theta)}$ is the Rayleigh excess ratio, which describes the ratio of the intensity at angle θ (I_θ) to that of the incident light (I_0) and can be measured by the light scattering detector. K is an experimental constant depending on the square of the solvent refractive index (n_0), the square of the refractive index increment (dn/dC) and the inverse fourth power of the incident wavelength (λ). The refractive index detector can determine n_0 and the concentration, C, of the protein while dn/dC has a constant value for all proteins, with M representing molecular mass of the protein. The non-ideality of the solution is accounted for by a power series about C, with B being the second virial coefficient. The low concentrations used for SEC-MALS and the dilution by SEC means the team 2BC becomes negligible. For large macromolecules with dimensions greater than λ/20 light can scatter from different parts of the molecule and cause destructive interference that reduces the intensity of scatted light. To correct for this the structure factor $P_{(\theta)}$ is required. This allows for extrapolation to θ = zero, where there is no destructive interference of the scattered light. To use the equation for mass determination further information about $P_{(\theta)}$ is required from theoretical analysis of large-particle scattering. Along with the relation of $P_{(\theta)}$ to the angular dependence of the scattered light intensity and the mean-square radius of gyration of a protein S. Replacing $P_{(\theta)}$ with this theoretical term a new equation can be obtained (Eq. 2.18-2).

Eq.2.18-2

$$\frac{KC}{R_\theta} = \frac{1}{M}\left(1 - \frac{16\pi^2}{3\lambda^2}\langle S^2\rangle \sin^2\left(\frac{\theta}{2}\right) + \ldots\right)$$

In SEC-MALS the differential refractive index detector measures the concentration of the sample with the MALS detector measuring simultaneously the light scattered at several angles. Using the equation above, a plot of KC/R_θ vs $\sin^2(\theta/2)$ is used to calculate the molecular weight from the intercept at $\sin^2(\theta/2)$ = zero and the gradient of the slope giving the radius of gyration (Harding and Jumel, 1998; Wyatt, 1993).

SEC-MALS was performed using the following setup. The ÄKTA purifier system with Superdex 200 10/300 analytical filtration column (GE healthcare) was connected to the DAWN® HELEOS™ 8 multi-angle light photometer (Wyatt technology), coupled to the Optilab® T-rEX™ refractometer (Wyatt

technology) in this specific order. The system was washed overnight at 0.5 mL/min in standard buffer to reduce noise in the scattering signal to <40 µV. For normalisation, 50 µL of MBP or BSA at 2 mg/mL was injected into the system prior to the protein samples for calibration of the inter-detector delay. Data were analysed using the ASTRA®6 software (Wyatt technology). The protein samples were diluted to ~2 mg/mL where necessary, with 50 µL of protein applied to the system. Data were collected using the ASTRA®6 software (Wyatt technology) with assistance from Dr Joseph Maman.

2.19 Pull-down experiments with CtIP and And1

For pull-down experiments 50 µL of amylose resin (NEB) was pre-equilibrated with pulldown buffer (25 mM HEPES pH 7.2, 100 mM KCl, 0.5 mM TCEP, 0.1% v/v igepal CA630) supplemented with 2 mg/mL BSA. Bait proteins (MBP-NTD and MBP-CTD) were incubated with the amylose resin for 1 h at 4°C with constant rolling. Amylose beads were pelleted by centrifugation at 1,000 RCF for 2 minutes at 4°C and washed twice with 500 µL of pulldown buffer. Prey proteins ($And1_{FL,}$ $And1_{316-1129}$, $And1_{826-1129}$, $And1_{1-316}$, $And1_{1-826}$, and $And1316_{-826}$) were buffer-exchanged using protein desalting spin columns (Cat no. 89862 Thermo fisher) into pulldown buffer supplemented with 2 mg/mL BSA. Following buffer exchange, prey proteins were incubated with the amylose beads, either with or without bait, for 1 h at 4°C with constant rolling. Amylose beads were pelleted by centrifugation at 1,000 RCF for 2 minutes at 4°C and washed twice with 500 µL of pull-down buffer A. MBP-bound protein was eluted from the beads with 30 µL of pulldown buffer supplemented with 20 mM maltose; the fractions were analysed by SDS-PAGE.

Where necessary, TEV digestion was performed using 3 µL of 2 mg/mL TEV added to 15 µL of bead eluate and incubated at room temperature for 60 minutes, shaking at 600 RPM.

2.20 Crystallisation and X-ray data collection

Initial crystallisation trials were performed using commercial 96 well crystallisation screen reagent kits (table 2.20). CtIP protein was thawed on ice prior to dilution and addition of glycerol where indicated. Where nucleation was an issue, protein was not centrifuged prior to dispensing.

For Strep-tagged constructs, optimal crystallisation conditions were at a final concentration of 7.5 mg/mL protein with 5% v/v glycerol added after protein thawing. Drop size and ratio was optimised for wild type strep tagged proteins to 400 nL of protein with 200 nL of reservoir. $NTD_{31\text{-}145}$DM-Strep gave crystals at a drop size and ratio of 200 nL protein + 200 nL of reservoir.

X-ray diffraction data were collected the MX beamlines IO3, IO4, IO4-1 and I24 at Diamond Light Source (DLS). During DLS maintenance the European Synchrotron Radiation Facility (ESRF) in Grenoble was used on the ID30B Micro focus beamline. Images were viewed on ADXV either on a Linux machine or a NoMachine connection to DLS on windows.

Crystallisation screen	**Company**	**Overview**
The Classics Suite I	Qiagen	A spectrum of the most popular chemicals in protein crystallography
JCSG+ ™	Qiagen	JCSG = Joint Centre for Structural Genomics Members of the JCSG identified “conditions most essential for promoting crystal formation for the most diverse set of proteins.”
pHClear I	Qiagen	Easy analysis of the effect of precipitant concentration with respect to pH
pHClear II	Qiagen	
PACT Suite	Qiagen	A screen looking at the effect of different Poly ethylene glycols with pH
MPD Suite	Qiagen	A combination of MPD with various salts, pH conditions and hits with MPD known to be successful for other proteins
Morpheus®	Molecular Dimensions	Derived from extensive data mining, uses a 3D grid covering a range of pH PEGs and salt additives with all conditions acting as a Cryogenic protective.
Wizard™ Classic 1 & 2	Molecular Dimensions	Random sparse matrix screen designed to increase the probability of producing crystals during initial screening of biological macromolecules.

Table 2.20 **List of commercial crystallography screens used.**

2.20.1 Crystallisation techniques

Conditions for optimisation were generated in accordance to the manufactures original condition. Each buffer condition was made as a “high” and “low” pH to enable generation of any pH in-between by dilution (table 2.20.1-1). 500 mL of each buffer was made at 0.5 M, pH corrected, filtered and

degassed all on the same pH meter sequentially and stored at 4°C. 50 mL of each condition was stored at 19°C ready for plate generation until used.

Buffer	pH	Supplier
MES	pH 5.5 Low	Melford Cat no. E24103
MES	pH 6.7 High	Melford
HEPES	pH 6.8 Low	Melford Cat No. H7530
HEPES	pH 8.2 High	Melford
TRIS	pH 7.0 Low	Sigma Cat no. T6791
TRIS	pH 9.0 High	Sigma
CHES	pH 8.6 Low	Sigma Cat no. 2885
CHES	pH 10.0 High	Sigma

Table 2.20.1-1: **Crystallisation stock buffers.**

Stock solutions of precipitants (table 2.20.1-2) were made in Milli Q water, filtered and degassed prior to storage at 4°C. In accordance to manufactures conditions they were not pH corrected. 50 mL of each condition was stored at 19°C ready for plate generation until used.

Precipitant	Volume made/ mL	Molecular mass/ g mol^{-1}	Concentration	Supplier
Ammonium sulphate	500	132.14	3M	Sigma Cat No. 31119
Sodium thiocyanate	50	81.07	2M	Sigma Cat No. 467871
Sodium tartrate	500	194.05	2M	Sigma Cat No. 228729
Sucrose	250	342.29	0.1M	Fisher Cat No. S/8600/60
Maltose	250	360.31	0.1M	Sigma cat No. 63418
Glycerol	250		50%	Sigma G5516-1L
Ethanol	50		35%	Sigma 02854-1L
MPD	50		35%	Sigma 112100
Methanol	50		35%	Sigma 603-001-00-X
Isopropanol	50		35%	Sigma 33539
SEC-butanol	50		20%	Sigma B-1888

Table 2.20.1-2: **Crystallisation stock precipitants.**

2.20.1.1 Manual screening

For manual screening 24 well XRL crystallisation plates were used (molecular dimensions). 1 mL of each 24 crystallisation conditions were made by manual pipetting with 0.5 mL dispensed into each well. Where hanging drop crystallisation method was used, protein drops were dispensed onto siliconized coverslips (Hampton research) with wells sealed with Dow Corning® vacuum grease. Sitting drop crystallisation used a Micro-bridge® (Hampton Research) in each well for a single protein drop. Plates were sealed with Duck Crystal clear tape.

2.20.1.2 Dragonfly® optimisation screens

Optimisation screens were designed using the Dragonfly® (TTP labtech) software package. Software was used to calculate dispense volumes for manual hand pipetted screens or CSV files for use with the Dragonfly® Robotics. The instrument was prepared with 19°C pre-equilibrated buffers and 96-well plates dispensed. Syringes and reservoirs were washed with Milli Q water before and after each use. Plates were mixed using the MXone® mixer (TTP Labtech) for 2 minutes prior to sealing with Duck Crystal Clear tape. MXone mixer was cleaned with Milli Q water before and after every use. Plates were incubated for a minimum of 20 minutes to equilibrate to room temperature before drop dispensing.

2.20.1.3 Seed Bead seed stock generation

For generation of a seed stock the Seed-bead™ (Hampton research) was used. Suitable crystals were selected and 5 µL of reservoir solution added to the drop, mixed, and aspirated into the seed bead tube on ice. 10 µL of reservoir solution was then washed around the crystal containing well and aspirated into the seed bead tube, this is repeated 8 times. Crystals were fragmented by a 30 second vortex mixing pulse with 1-minute interval on ice for a total of 2 minutes.

2.20.1.4 Seed Bead Seed stock from a single crystal

Crystal was thawed directly into an equilibrated crystallisation condition, identical to the original condition, for aspiration and seed generation by seed bead. The crystal loop was removed from liquid nitrogen and placed directly into a clear 400 nL equilibrated crystal drop, 5 µL of reservoir solution was added and aspirated into a seed bead tube. A further 50 µL of reservoir solution was aspirated, 10 µL at a time into the well and then into the seed bead tube. Seeds were generated by vortex mixing using a 20 seconds pulse with 1-minute interval on ice for a total of 2 minutes. For Mosquito® seeding 10 nL of undiluted seeds were mixed with 190 nL of reservoir solution prior to addition to protein drops.

2.20.1.5 Manual streak seeding

For manual seeding the seed stock was used undiluted. A seeding whisker (Hampton research) was used to drag through three consecutive identical protein crystal drops to give a dilution of seeds. For hanging drop crystallisation three drops were dispensed onto the same coverslip, for sitting drop crystallisation either in 96 or 24 well format three adjacent wells were used.

2.20.1.6 Mosquito® micro seeding

The identical condition from which the seeds were made from was produced to enable serial dilutions from the concentration seed stock of 1:10, 1:100, and 1:1000. The Mosquito® Robottics (TTP labtech) multi-aspirate mode was used to mix 5% seed with 95% reservoir condition, for example, 10 nL of seeds with 190 nL of reservoir solution prior to dispensing into the drop. Optimum seed concentration is determined experimentally with fresh dilutions generated from the stock on each preparation.

2.20.2 Crystallisation of NTD_{31-145}DM-Strep

The Morpheus® (Molecular Dimensions) crystal screen was performed with purified NTD_{31-145}DM-Strep protein in buffer containing 20 mM Tris pH 8.0, 300 mM NaCl, 5% v/v glycerol at 7.5 mg/mL. The protein was mixed with crystallisation buffer and micro seeds at a 1:19:1 ratio in 0.4µL in a 96 well MRC 2 drop crystallisation plate using the Mosquito® robot. Final crystallisation condition for structure determination was G6 (MDSR-47-2-30), containing 30% w/v EDO_P8K (40% v/v ethylene glycol; 20 % w/v PEG 8000), 0.1M MB2 pH 7.5 (sodium HEPES; MOPS (acid)), 10% MCA (0.2 M Sodium formate; 0.2 M ammonium acetate; 0.2 M sodium citrate tribasic dihydrate; 0.2 M sodium potassium tartrate tetrahydrate; 0.2 M sodium oxamate).

2.20.3 Visualisation of protein models.

Determined protein crystal structures and SAXS models were visualised using PyMol. The “align” tool was used to align the NTD_{31-145}DM-Strep structure with PDB entry 4D2H to generate a model of CtIP residues 18-136. This model was further overlaid with the NTD_{18-145} model to fit the SEC-SAXS data manually. Further SEC-SAXS models described in the main text were overlaid manually. The model to fit the SEC-SAXS data for MBP was overlaid with the known crystal structure 3HPI manually.

Chapter 3, CtIP N-terminal domain

3.1 Introduction

The N-terminal domain (NTD) of CtIP possesses conserved structural elements that have been shown to be important for CtIP activity (Davies et al., 2015). Functional studies of CtIP have illustrated its critical role in homologous recombination (HR) mediated DSB repair, with deletion studies of the N-terminal region showing CtIP-null like results highlighting its importance. U2OS cells expressing CtIP with an N-terminal 200 amino acid deletion showed sensitivity to Ionising radiation and camptothecin in EGFP-based HR repair assays, with complete loss of self-association between CtIP alleles (Wang et al., 2012). In IR treatment experiments, HeLa cell lines possessing an N-terminal deletion of 17-160 exhibited loss of Chk1 phosphorylation, reduction in RPA DNA recruitment and reduced DSB repair efficiency through HR (Yuan and Chen, 2009). In immunoprecipitation studies an N-terminal domain of CtIP comprising residues 1-160 was sufficient to interact with the MRN (mre11-Rad50-Nbs1) complex, with deletion of residues 22-45 removing the interaction with Nbs1 (Yuan and Chen, 2009). In contrast, mutation of the central tetramerization Lysine on CtIP, resulting in dimeric CtIP, had no effect on MRN interaction (Davies et al., 2015).

Experimentally derived models of the tertiary CtIP structure are limited to short protein interaction sites (figure 1.6-3 introduction) and the putative coiled-coil region at the N-terminus (Fusco et al., 1998). The latter region covers residues 45-160, and early studies using analytical ultracentrifugation (AUC) demonstrated that it could self-associate *in vitro* (Dubin et al., 2004). This domain could be further divided into two sub domains: CtIP residues 45-92 and CtIP residues 93-160. Both were seen to form homodimers with some coiled-coil characteristics (Dubin et al., 2004). More recently, residues 18-52 of CtIP were observed to form a tetrameric species, which was dependent on a lysine at position 27 (Davies et al., 2015). Tetramer formation was also found to be essential for correct functioning of CtIP *in vivo* (Davies et al., 2015).

To date, these 34 residues compose the largest functionally relevant domain of CtIP that has been structurally characterised. While the architecture of the coiled-coil domain (residues 52-160) is unknown, structural and biophysical analysis of this region should clarify whether CtIP is largely elongated or compact. To further explore the N-terminal region of CtIP, domains contained therein were cloned for biophysical and crystallographic characterisation.

3.2 Results

3.2.1 Expression, purification and crystallisation trials of the NTD

Tetramerisation of the NTD has been shown by crystallography to have a dimer of dimers orientation (Davies et al., 2015). However, residues 18-52 comprise the largest determined structure of CtIP to date. To further our understanding of the domain architecture a structure of residues 18-145 (NTD_{18-145}) was desired including the tetramerisation and dimerisation domains. Initial crystallisation of NTD_{18-145} gave no tractable conditions. Crystallisation experiments were therefore focused specifically on the coiled-coil domain to enable an ensemble of the two structures to be generated.

As leucine 27 is the central residue in tetramerisation (Davies et al., 2015), the first construct spanned from CtIP residue 28 to 145 (NTD_{28-145}), including the zinc binding site at C89 and C92 (CXXC), (figure 2.2.1). The C-terminal boundary of this domain at 145 was previously determined by colleagues (Sun, 2013). The DNA was cloned from the CtIP image clone (4830693) into pMAT11 to generate an N-terminal His_6 MBP tagged fusion protein.

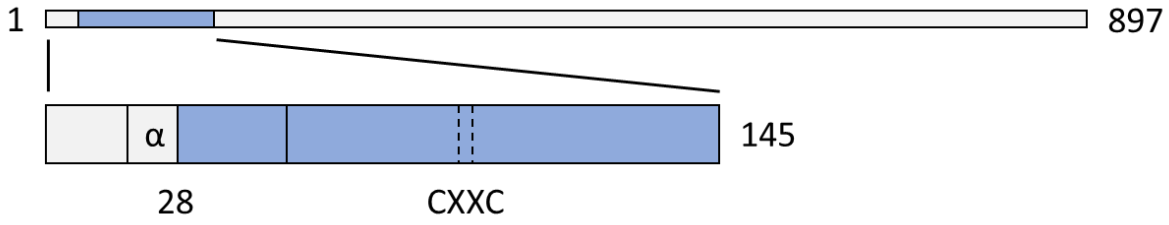


Figure 3.2.1: **Schematic drawing of NTD_{28-145}.** CtIP residues 28-145 are highlighted in blue, with the alpha helical tetramerisation domain (α) and zinc binding domains shown (CXXC).

3.2.1.1 Recombinant protein expression and purification

Protein expression was performed in Rosetta™ 2 (DE3) using sequential Ni-NTA, Amylose chromatography, on-column TEV cleavage and a final Ni-NTA recapture. Final polishing was through Superdex 75 16/60 SEC. As NTD_{28-145} contains no tryptophan residues, and only three phenylalanine residues, the absorbance at 280 nm was minimal with fractions analysed by SDS-PAGE to identify NTD_{28-145} (figure 3.2.1.1). Fractions corresponding to NTD_{28-145} were pooled for further analysis. Purifications of NTD_{28-145} contained a co-migrating higher molecular weight band, MALDI-TOF mass-spectrometry (PNAC facility, Department of Biochemistry, University of Cambridge) confirmed this to contain NTD_{28-145} and therefore a SDS-PAGE gel artefact. Multiple purifications of NTD_{28-145} were used to calculate the protein concentration and for crystallisation.

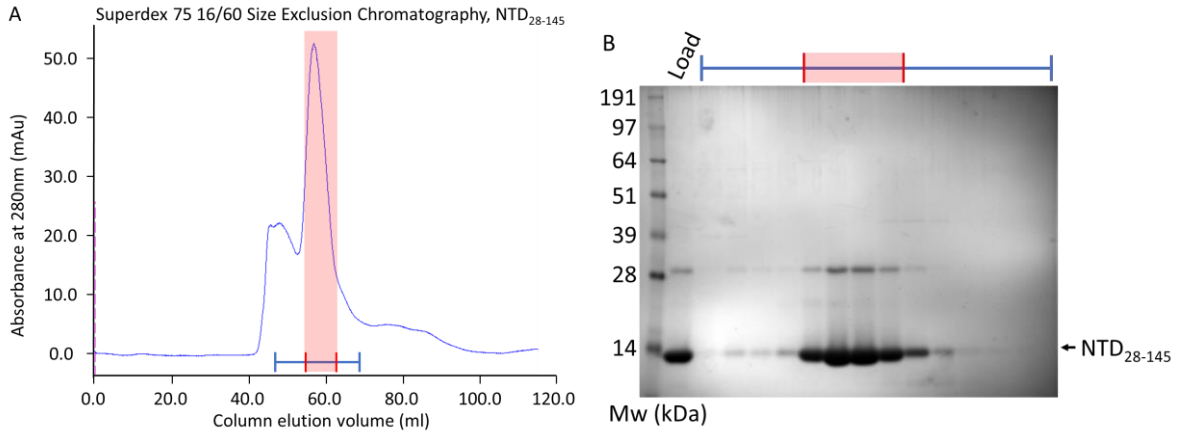


Figure 3.2.1.1: **Purification of NTD_{28-145}. A)** Chromatogram of the Superdex S75 16/60 SEC with fractions analysed by SDS-PAGE in B. **B)** final Ni-NTA recapture flow-through (Load) which was concentrated for SEC. Blue lines indicate fractions analysed by SDS-PAGE from A, with the red box indicating pooled fractions for concentration and freezing.

3.2.1.2 Protein concentration determination of the coiled coil domain

For systematic screening of crystal conditions an accurate protein concentration was required. The Bradford assay was used, however the linear range of NTD_{28-145} was below 0.35 mg/mL. Desired concentrations for crystallography were between 5-15 mg/mL, dilution and extrapolation from the linear range for the Bradford assay would introduce a wide error range in concentration predictions and affect reproducibility. Thus, Amino Acid Analysis (AAA) was used to determine protein concentration.

3.2.1.2.1 Amino acid analysis and generation of a standard curve

Each individual amino acid contributes to the protein concentration in AAA giving a more accurate result. This process was carried out by Mr Peter Sharratt in the Department of Biochemistry PNAC facility (University of Cambridge). To overcome the need to analyse every sample, a standard curve was generated correlating the AAA result to the weak NTD_{28-145} UV absorbance measured at 280 nm. Purifications of NTD_{28-145} were analysed by AAA obtaining 10 AAA results from 5 independent purifications. Measurements were also taken for the weak 280nm absorbance with a linear relationship between the two measurements up to 15 mg/ml as recorded by AAA. After 16 mg/mL the correlation between the readings deteriorate, however the linear region is sufficient to calculate protein concentrations for crystallography. Measurements were taken at 230 nm for the peptide backbone absorbance however there were no correlations. The standard curve equation Y =6.814x +0.2354 (where X is absorbance at 280 nm at a 1:1 ratio and Y is mg/mL) was used for calculation of protein concentration for all NTD constructs without a tryptophan residue.

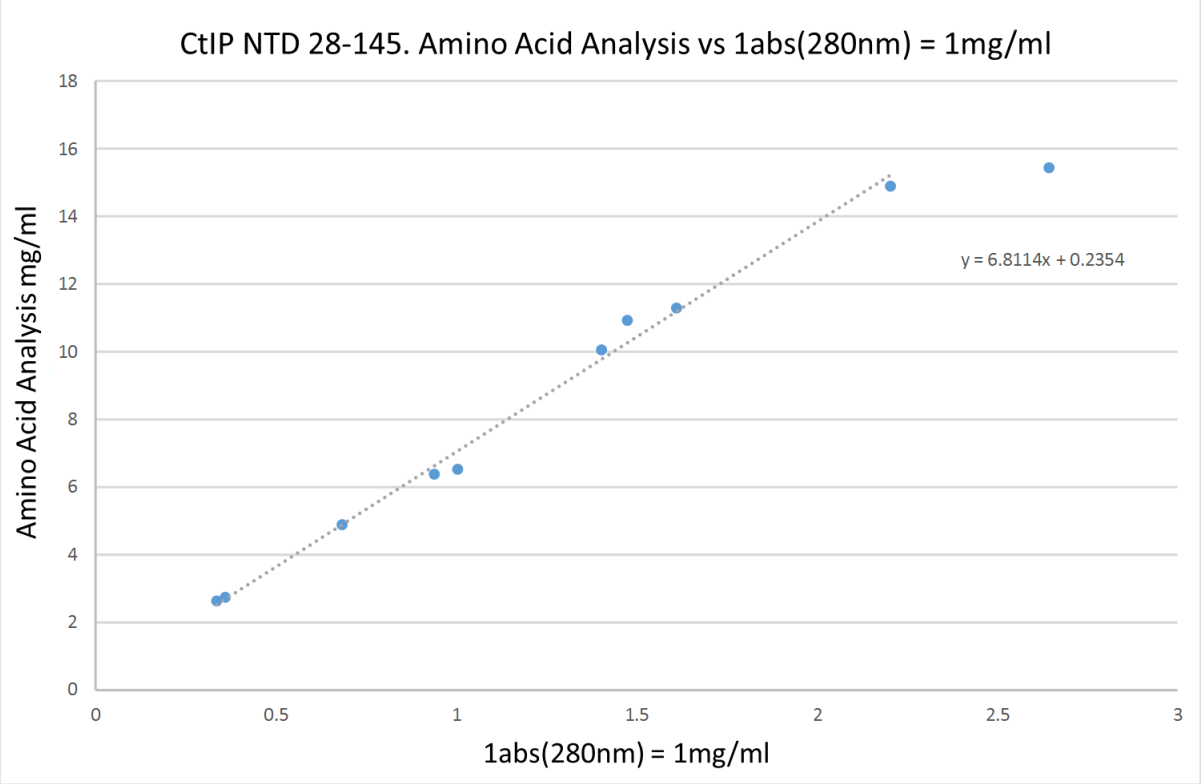


Figure 3.2.1.2.1: **Amino Acid Analysis standard curve of NTD_{28-145}.** Protein absorbance at 280 nm at a 1 absorbance unit to 1 mg/mL ratio was plotted against absolute protein concentration measured by Amino Acid Analysis. A line of best bit was calculated using data within the linear range. R^2 for the linear relationship is 0.9927.

3.2.2 Crystallisation of the N-terminal coiled coil domain

The purified NTD_{28-145} was dispensed into commercially available crystallisation conditions: Morpheus ®, Wizard™ 1&2, JCSG+™ pHClear 1 and pHClear 2 using the Mosquito® crystallisation robot for vapour diffusion at 7.5 mg/mL. Only small nuclei were seen that could be either salt or protein crystals. Optimisation of these conditions by reducing the salt concentration increased nucleation, while addition of zinc ions failed to show any crystals. Manual pipetted screens were dispensed using the hanging-drop method, although nuclei were present there were no signs of crystal growth. After 20 crystallisation trials there were no defined crystals to optimise.

3.2.2.1 Thermal shift analysis of $NTD_{28\text{-}145}$

The analysis of a protein thermal melt by binding to SYPRO® orange can be used to determine the optimum buffer for protein stability (Huynh and Partch, 2015). The higher the melt temperature required the more stable the protein is for a given buffered condition. Melt curves for $NTD_{28\text{-}145}$ were generated using a buffer screen generated by the Hyvönen group (Appendix figure 7.1). From the 96 conditions many gave no point of inflection with reductions in stability, the top 4 hits are shown in figure 3.2.2.1. The standard buffer gave the greatest melt peak of 47°C, showing the protein buffer was optimal for stability. Construct optimisation was performed next to encourage crystallisation.

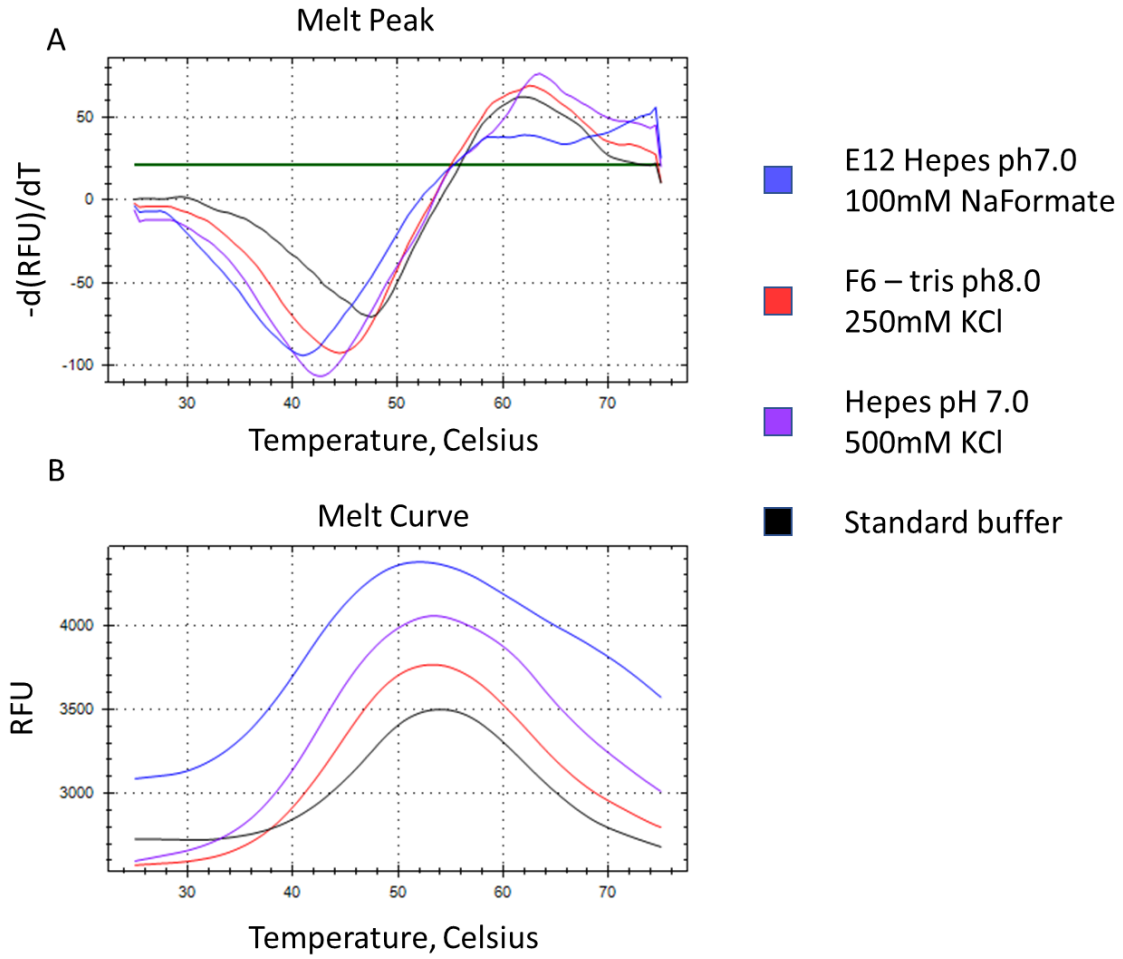


Figure 3.2.2.1: **$NTD_{28\text{-}145}$ thermal melt peak and melt curve.** From the binding of SYPRO orange to unfolded protein the fluorescence increases and is measured in relative fluorescence units.

3.2.2.2 Construct optimisation and protein purification.

For construct optimisation the N-terminal start residue was targeted, changing the position by one amino acid per construct generating 10 start positions from residue 29 to residue 38. These were cloned into pMAT11 to generate His_6-MBP tagged constructs (figure 3.2.2.2-1).

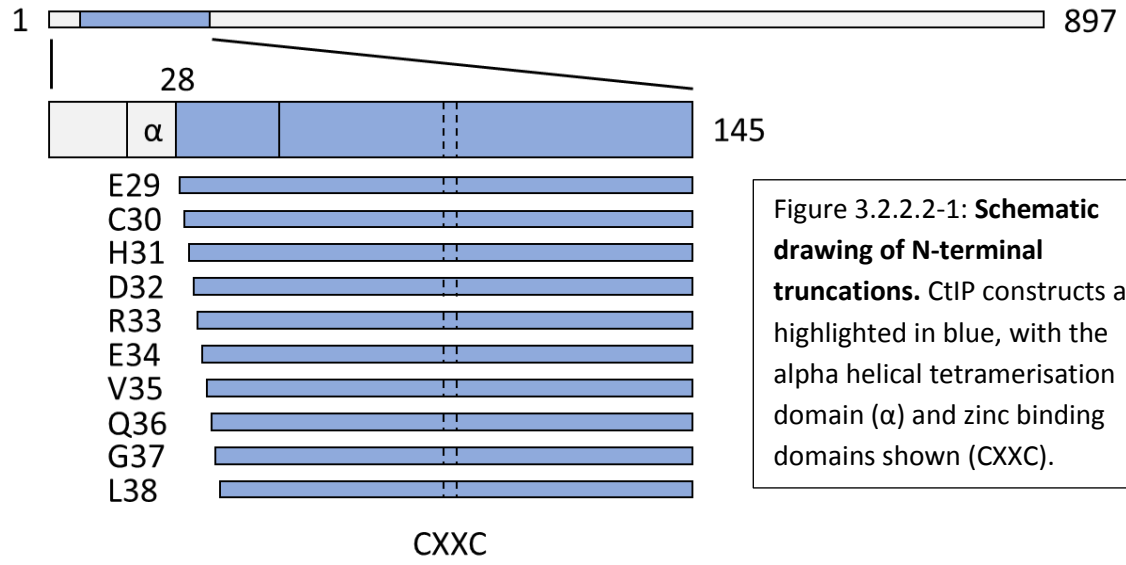


Figure 3.2.2.2-1: **Schematic drawing of N-terminal truncations.** CtIP constructs are highlighted in blue, with the alpha helical tetramerisation domain (α) and zinc binding domains shown (CXXC).

The new constructs were expressed in Rosetta™ 2 (DE3) and purified using the Ni-NTA purification system, with soluble protein expressed for all constructs. Figure 3.2.2.2-2 is an example of an original (a/b) and optimised (c/d/e) NTD_{31-145} purification, with identical optimised purifications for the other protein variants. For both purifications the soluble protein was bound to Ni-NTA resin, washed and eluted. Protein was subsequently dialysed overnight with 2 mM DTT and addition of TEV removing the His_6MBP tag from the NTD proteins. Unfortunately, the reaction did not go to completion generating three stable products (figure 3.2.2.2-2B/C). The sizes these products correspond to the molecular mass of NTD_{31-145} (14.2 kDa), the His_6-MBP tag (44.7 kDa) and the un-cleaved fusion protein (58.9 kDa.) TEV protease (27 kDa) was also detectable by SDS-PAGE. Desalting chromatography removed the DTT from the protein buffer to enable recapture by Ni-NTA. Figure 3.2.2.2-2C shows sequential flow-through through Ni-NTA, Amylose and Ni-NTA recapture columns, with removal of contaminants at each stage by binding to the column allowing the NTD to flow through. In the optimised purification a third Ni-NTA stage removed further contamination. The Ni-NTA#2/3 flow through protein was subsequently concentrated and applied to a Superdex 75 16/60 SEC. SDS-PAGE analysis showed the optimisations removed the vast amount of contamination (figure 3.2.2.2-2 A/B and D/E) at the cost of NTD_{31-145} yield from the addition of the Ni-NTA stage prior to Amylose re-capture. A protein concentration dependency was also apparent for the SDS-PAGE 'dimer' band. The very weak 280 nm absorbance made tracking the protein by SEC difficult (figure 3.2.2.2-2D) with an intense band by SDS-PAGE giving

only a 10 mAu absorbance reading on the AKTA purifier at 4°C. Protein concentration was however determined using the NTD_{28-145} AAA calibration curve for all purifications.

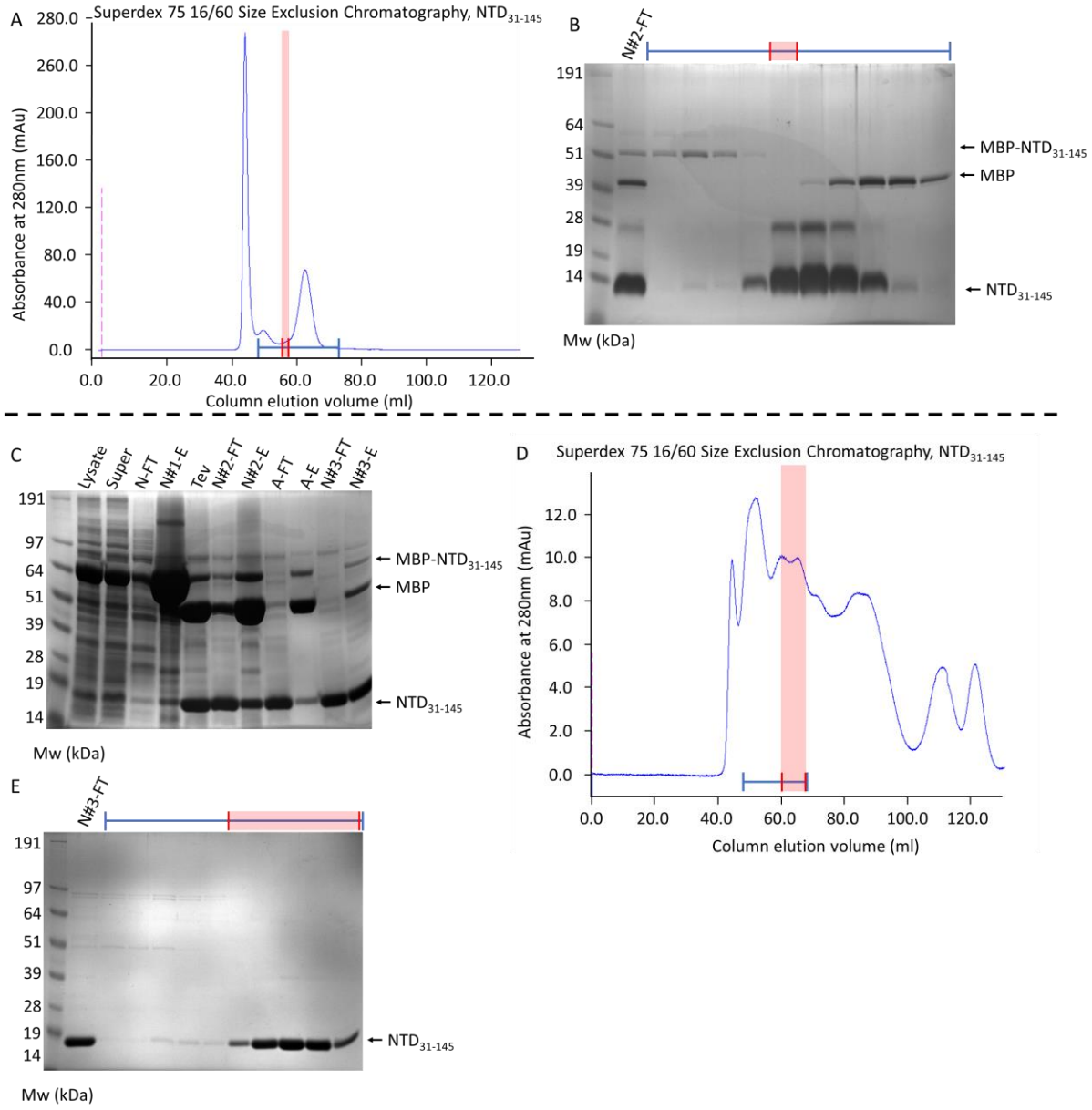


Figure 3.2.2.2-2**: Purification of NTD_{31-145} by Ni-NTA and amylose chromatography. A)** Original NTD_{31-145} final stage purification Superdex 75 16/60 SEC chromatogram. **B)** SDS-PAGE analysis from SEC fraction, only one fraction was pooled for analysis. The dotted line separates original from the optimised purification. **C)** Protein was purified by Ni-NTA (N) with elution factions (N#1-E) dialysed overnight with TEV protease to remove the tags. Protein was further purified using an addition Ni-NTA column (N#2) with the flow-through (FT) passed over an amylose (A) column. This amylose flow-through (A-FT) was further passed over a 3rd Ni-NTA column with flow-through (N#3-FT) concentrated for SEC in D. **D)** Chromatogram of the Superdex 75 16/60 SEC. Blue lines indicate fractions analysed by SDS-PAGE in E. **E)** SDS-PAGE analysis shows a large reduction in contaminations at the cost of the protein yield. Fractions indicated in red were pooled and concentrated for analysis.

3.2.2.3 Initial screening of N-terminal truncations

Constructs NTD_{29-145}, NTD_{30-145}, NTD_{31-145}, NTD_{32-145}, NTD_{33-145}, and NTD_{38-145} were purified identically and dispensed using the Mosquito® crystallisation robot at 7.5 mg/mL into the commercially available Morpheus®, Wizard™ 1&2 and pHClear I screens. NTD_{31-145} was the only construct to show signs of crystallisation in Wizard™ 1&2; one condition (D6) containing 15% v/v ethanol and 100 mM Tris pH 7.0 gave a shower of small needle crystals (figure 3.2.2.3.) The same Wizard 1&2 screen was repeated with two concentrations of NTD_{31-145} with the high concentration giving the same needle crystals. Lower concentrations of NTD_{31-145} however failed to crystallise (figure 3.2.2.3 C/D.)

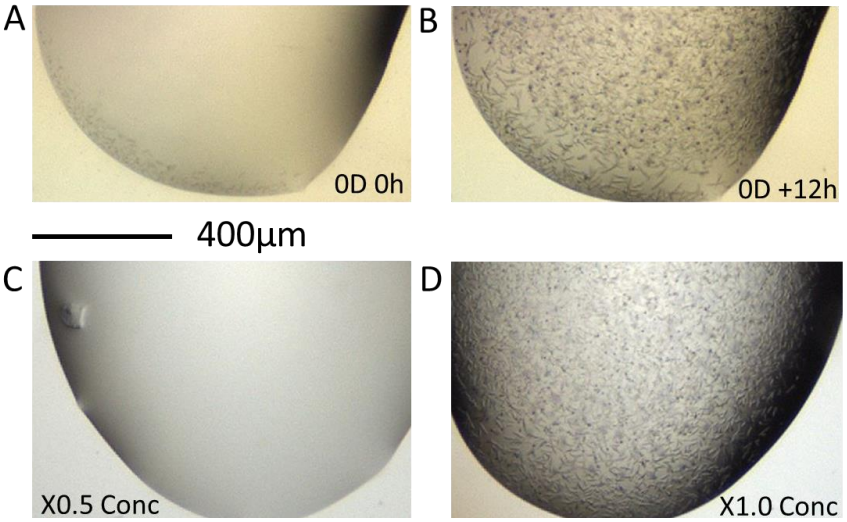


Figure 3.2.2.3: **Wizard 1&2 condition D6, NTD_{31-145} Needle crystal.** X1.0 is the same 7.5 mg/mL concentration as the previous screen, and X0.5 is half protein concentration. Both images C/D were taken at the same time.

3.2.2.4 Optimisation of NTD_{31-145} Needle crystals

Hand pipetted screens failed to reproduce crystallisation. A 96-condition screen was made using the Dragonfly® discovery robot varying ethanol concentration and Tris pH. Using the Mosquito® robot protein was dispensed with larger crystals at pH 7.2 and lower ethanol concentrations (figure 3.2.2.4-1).

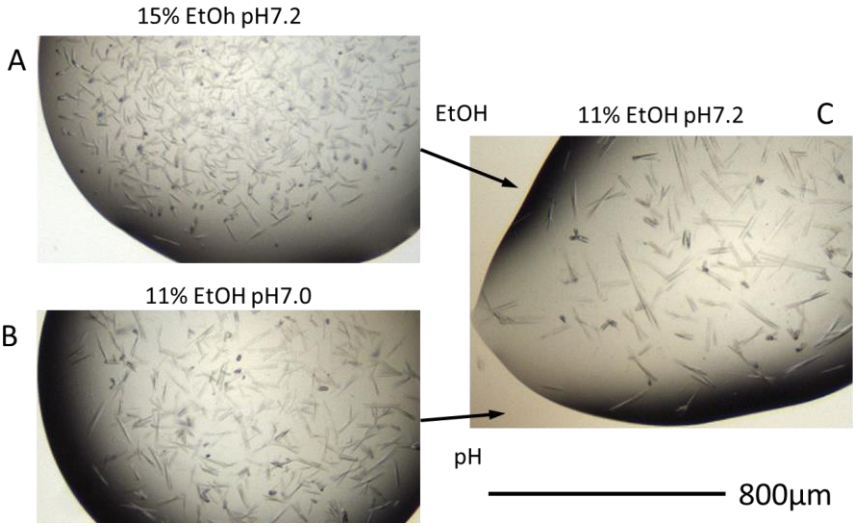


Figure 3.2.2.4-1: **Effect of Ethanol and pH concentration on NTD_{31-145} crystal formation**. Dragonfly® optimisation screen with varying ethanol concentration on the X axis from 10-21% and the Tris pH on the Y-axis from pH 6.8 to 8.8.

Crystals were growing mainly in one direction with a high density. To further optimise the crystals the protein to crystallisation condition drop ratio was varied from 1:1, 2:3 and 1:2. Ethanol concentrations (10-17%) and Tris pH (7.0-7.3) were further varied. However, batch to batch protein variations meant only one drop at a ratio of 1:2 contained a needle crystal (figure 3.2.2.4-2A). Protein concentration was assessed using the 1:2 ratio. Panels A and B of figure 3.2.2.4-2 show the same condition and protein concentration between the two screens with vast differences in nucleation and crystal growth. Reducing the protein concentration gave a reduction in crystal size with vast amounts of nuclei seen. Surprisingly, the small changes in pH and ethanol which from this screen gave drastic changes in number and appearance of crystals (figure 3.2.2.4-2 B/C/D).

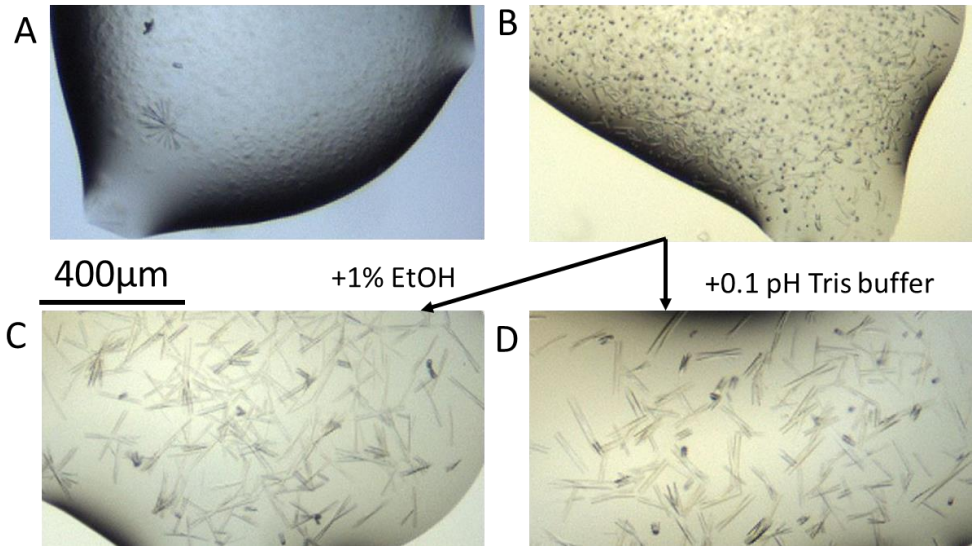


Figure 3.2.2.4-2: **NTD_{31-145} ethanol condition optimisation**. The ethanol concentration was varied from 8-15% with Tris pH varied from 7.2-7.4. Small changes in concentration changed the appearance of the needle crystal with a drastic plate to plate variation.

Reproducibility of the needle crystals was poor. Crystals were also fragile which impeded crystal mounting for x-ray diffraction. Two crystals were mounted onto cryo loops by Dr Paul Brear (University of Cambridge X-Ray Facility) and exposed to synchrotron radiation at ESRF with no diffraction.

In an attempt to change crystal form four other alcohols were used: isopropanol, methanol, MPD and 1-butanol. From this optimisation only, small nuclei were seen in a single drop with large crystals absent from the control 11% v/v ethanol, tris pH 7.2 condition. To try to recapitulate the large needle crystals, a commercial MPD screen, and further optimisation screens were performed with different protein concentrations, fresh alcohols, and fresh purifications with the crystals seen in figure 3.2.2.4-2D unable to be reproduced. After a total of 9 NTD_{31-145} purifications and 27 crystallisation trials a new approach was necessary.

3.2.3 Crystallisation optimisation of the N-terminal coiled-coil domain

Many of the complications from the first rounds of crystallisation were due to uncertainty in protein concentration and intrinsic affinity of the NTD to Ni-NTA resin, with a large amount of NTD protein retained on Ni-NTA recapture. These were combined with the small window for crystal formation and volatility of alcohol conditions. To address these issues, the construct design and purification strategy was altered to include a C-terminal Strep affinity tag to allow recapture of the NTD protein following cleavage to increase purity. With the short sequence and tryptophan residue, the Strep II Tag along with Strep-Tactin resin was perfect for this application as the tryptophan also gave an absorbance at 280 nm.

3.2.3.1 Construct design

For addition of the C-terminal tag constructs from $NTD_{28\text{-}145}$ through to $NTD_{38\text{-}145}$ (figure 3.2.3.1) were cloned into pMAT11 with the WSHPQFEK (Strep tag) protein sequence added in frame at the c-terminus of the protein by primer extension. This generated a His_6-MBP-NTD-Strep fusion protein. All constructs were cloned using the same method and were sequence-verified.

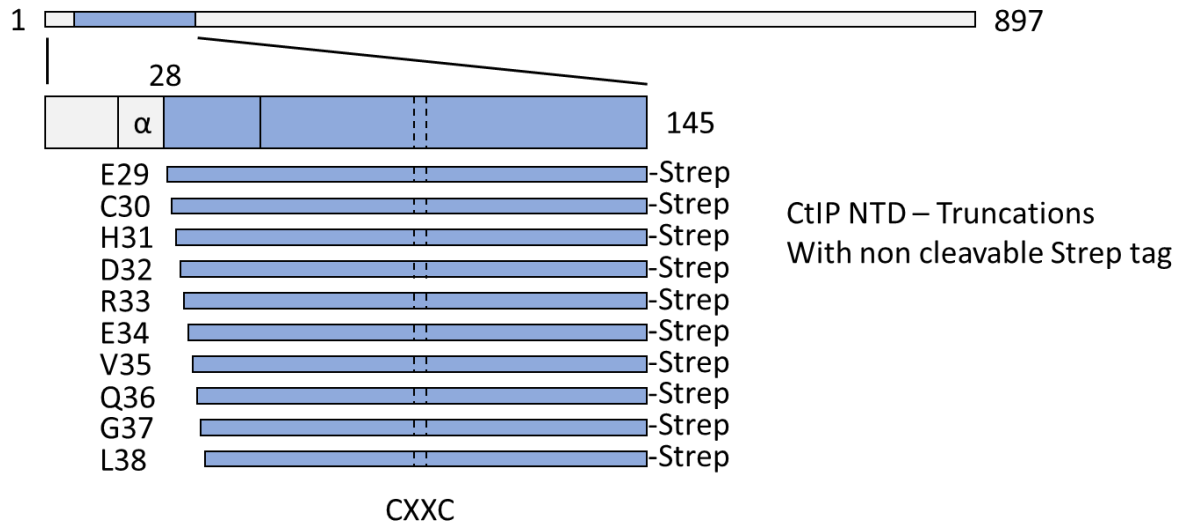


Figure 3.2.3.1: **Schematic drawing of C-terminal Strep tagged NTD constructs.** CtIP constructs are highlighted in blue, with the alpha helical tetramerisation domain (α) and zinc binding domains shown (CXXC).

3.2.3.2 Recombinant protein expression and purification

Protein expression was performed in Rosetta™ 2 (DE3) cells with all protein variants expressed and purified using the same techniques. Protein was soluble following lysis and centrifugation, and was subsequently purified using Ni-NTA affinity chromatography utilizing a 10 mM imidazole wash to remove weakly-bound impurities prior to final elution. Protein was dialysed overnight with addition of TEV to cleave the MBP tag producing a mixture of cleaved and un-cleaved products. The strep tag affinity to Strep-Tactin resin captured the NTD with flow rate reduced to 0.1-0.2 mL/minute. After washing with standard buffer, the protein was eluted from the Strep-Tactin resin and bound to a Hi-trap Ni-NTA column using the intrinsic affinity of the NTD. Figure 3.2.3.2A/B shows this small affinity with a step to 6% buffer B removing some contaminants and 10% buffer B removing the NTD protein. The His_6MBP retains bound to the column due to its high affinity, allowing removal of 99% of the contaminants. The combined fractions containing the NTD-Strep protein were applied to Superdex 75 16/60 SEC with a sharp symmetrical peak (figure 3.2.3.2C). Compared with non-strep tagged NTD constructs, the addition of the c-terminal Strep tag enabled higher purity without excessive use of Ni-NTA chromatography (figure 3.2.3.2D) and allowed the purification to be followed by 280 nm absorbance. Further to this, the strep tag enabled reliable protein concentration determination due to the presence of a tryptophan residue in the tag sequence with concentrations between 10-15 mg/mL achieved prior to aliquoting and freezing. The higher molecular weight “dimer” band was present for strep-tagged proteins for all the 10 different NTD-Strep constructs.

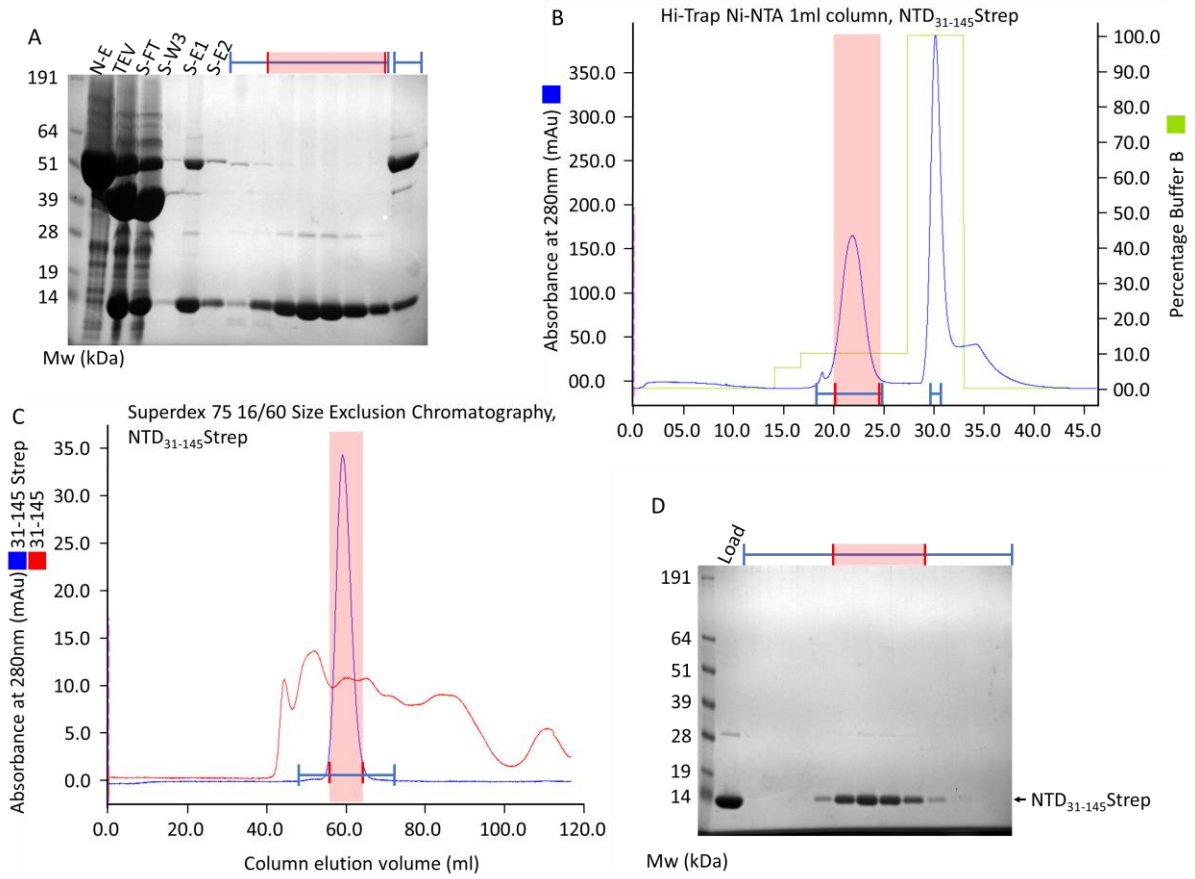


Figure 3.2.3.2**: Purification of C-terminal Strep tagged NTD construct NTD$_{31-145}$Strep, using Ni-NTA and Strep-Tactin chromatography. A)** SDS-PAGE analysis of stages after initial capture. Protein was initially purified through Ni-NTA with elution fraction (N-E) subject to TEV cleavage (TEV) to remove the MBP tag. Protein was bound to a Strep-Tactin column with flow-through (S-FT), wash (S-W) and elution (S-E) fractions shown. Elution 1&2 were pooled for subsequent His-trap Ni-NTA. Blue lines indicate the fractions analysed from B, with fractions highlighted in red pooled for SEC in C. **B)** Chromatogram for the His-Trap Ni-NTA, intrinsic NTD$_{31-145}$Strep affinity to the column allowed purification eluting with 10% imidazole. **C)** Blue trace – SEC Superdex 75 16/60 of fractions indicated in red from A/B for NTD$_{31-145}$Strep. Red trace – comparable SEC analysis of NTD$_{31-145}$. Both traces are independent runs on the same column, absorbance values are comparable. **D)** SDS-PAGE of the fraction loaded onto the SEC column (load) prior to analysis of SEC elution fractions. Fractions containing NTD$_{31-145}$Strep (red highlight) were pooled for concentration to 10-15 mg/mL and freezing in liquid nitrogen.

3.2.3.3 Initial screening of N-terminal truncations

The NTD-strep constructs starting from residues 30, 31, 32, 33, 35 and 38 were dispensed into the Morpheus®, Wizard™ 1&2, pHClear and JCSG+™ crystal screens. Only NTD_{31-145}Strep and NTD_{32-145}Strep in Wizard™ 1&2 gave crystals. Figure 3.2.3.3 shows these hits and any differences between the two constructs. Needles along with a further pyramid crystal form were present, appearing in ammonium sulphate, sodium tartrate and ammonium phosphate conditions. For crystal condition G5 the single amino acid difference changed the crystal form from needles to pyramids (figure 3.2.3.3).

There were no large differences between ammonium sulphate and sodium tartrate conditions. With the addition of the Strep-tag UV absorbance now could be used to indicate NTD-Strep crystals. The UV detection shows immediately that the crystal in G9 did not contain NTD-Strep. Although the pHClear screen gave 0 hits, the crystals formed over a large range of pH from pH 6.0 MES to pH 9.5 CHES. The initial Wizard™ 1&2 screen was repeated confirming the presence of medium sized crystals with well-defined edges. The larger pyramid crystals were mounted on crystal loops for X-ray diffraction without addition of cryo-protectant. Large ice rings were apparent however diffraction maxima were present at 36 Å. Now that crystals were appearing reproducibly, and had the potential to diffract, necessary optimisation was carried out.

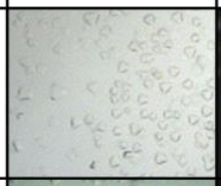
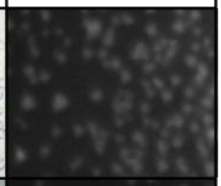
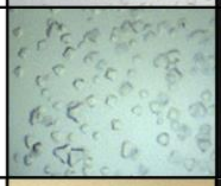
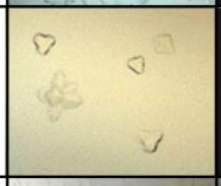
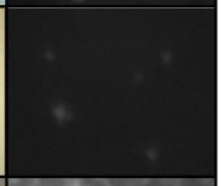
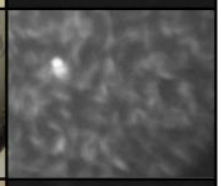
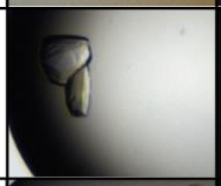
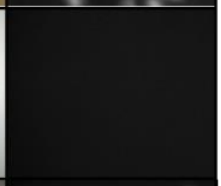
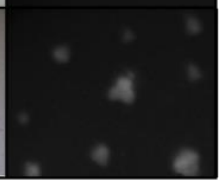
Protein	Wizard 1& 2 condition	Crystal image	UV image
NTD_{31-145} Strep	B6 1M K Na Tart 0.1M Imidazole pH8 0.2M NaCl		
NTD_{31-145} Strep	D11 1.26M $(NH4)_2SO_4$ 0.1M TRIS pH 8.5 0.2M $LiSO_4$		
NTD_{31-145} Strep	F3 1.26M $(NH4)_2SO_4$ 0.1M HEPES pH 7.5		
NTD_{31-145} Strep	G5 1.26M $(NH4)_2SO_4$ 0.1M CHES pH 9.5 0.2M NaCl		
NTD_{32-145} Strep			
NTD_{31-145} Strep	G9 1M $(NH4)_2HPO_4$ 0.1M Na_3 Citrate pH 5.5 0.2M NaCL		
NTD_{31-145} Strep	H9 1.26M $(NH4)_2SO_4$ 0.1M MES pH 6		

Figure 3.2.3.3: **Wizard™ 1&2 screening of strep tagged NTD constructs.** Where no differences were seen the best crystal image was selected.

3.2.4 Optimisation of NTD_{31-145}Strep

The first optimisations were carried out using a 2D grid screen of ammonium sulphate vs buffer pH prepared using the Dragonfly® crystallisation robotics. See chapter 2 for stock solutions and preparation. Figure 3.2.4-1 shows the first optimisation grid screen and the appearance of crystals within the screen, three crystal forms were present: pyramids, needles and rods each with their own distribution over the plate. Pyramids could be seen in the presence of rods, along with pyramids and rods appearing adjacent in similar conditions showing the sensitivity of crystal form formation.

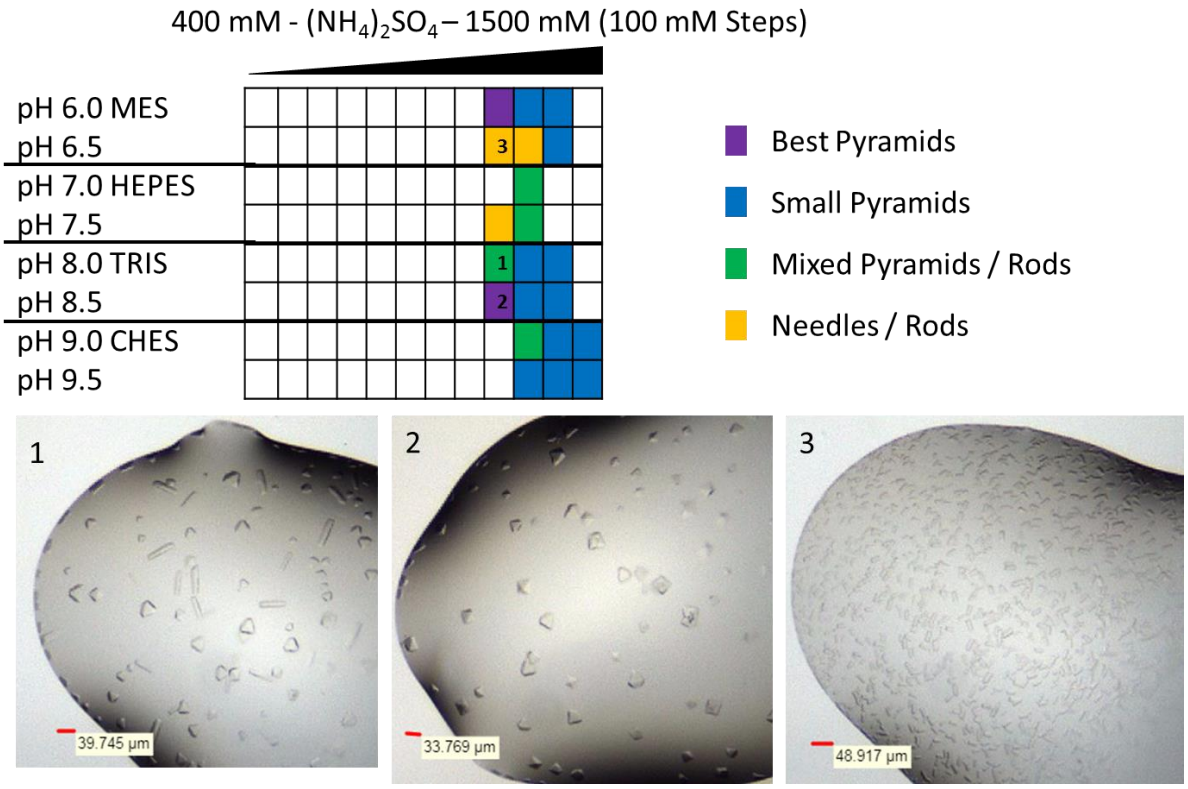


Figure 3.2.4-1**: Optimisation 1 of NTD_{31-145}Strep crystals.** Top panel shows the Dragonfly® optimisation screen with a range of ammonium sulphate concentrations over the X axis for 12 wells. Colours indicate crystal form shown in legend with blank squares having 0 crystals. Numbers correspond to the bottom panel crystal images.

Data collection without cryo-protection gave large ice rings with minimal diffraction. Addition of a 20% v/v glycerol, 80% v/v reservoir mixture directly to the drop prior to crystal mounting and freezing has been seen to provide cryo protection. Without a low percentage of cryo-protectant within the crystallisation condition this can cause the crystal to crack which is detrimental to diffraction (Holzer, 2017). To generate divergence between the crystal forms, enable cryo-protection and increase crystal size for the second optimisation: drop size was increased while maintaining ratio, the range of

ammonium sulphate concentrations reduced, and 2% v/v glycerol added to half of the plate to enable cryo-protection.

The larger 600 nL reservoir mixture (figure 3.2.4-2) in some cases doubled crystal size with a reduction in mixed crystal form drops. The addition of 2% v/v glycerol increased the requirement for ammonium sulphate (figure 3.2.4.2-2 – A11/A12), however the collective changes allowed assessment by X-ray diffraction. The crystals highlighted by the measurements in figure 3.2.4-2 were mounted on cryo loops. For this procedure 2 µl of a solution containing 20% v/v glycerol 80% v/v reservoir solution was added to each crystal drop and the crystal mounted with no signs of reduced stability. Crystals were tested for diffraction at the Diamond Light Source synchrotron on IO4. Only one crystal (A11) gave Bragg maxima at 36 Å, however cryo-protection was successful in removing ice rings.

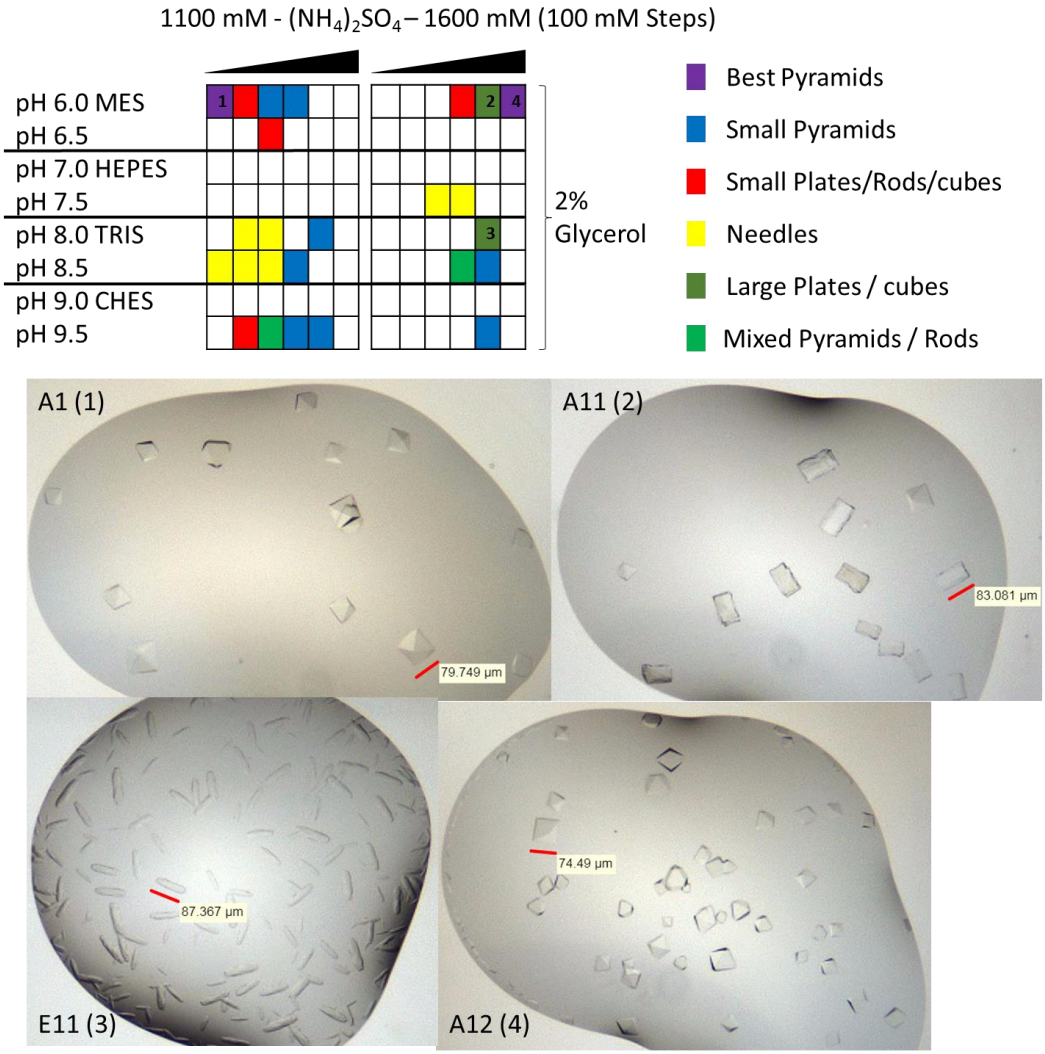


Figure 3.2.4-2**: Optimisation 2 of NTD_{31-145}Strep crystals.** Top panel shows the optimisation screen with a range of ammonium sulphate concentrations over the X axis for 6 wells repeated on the Left-Hand Side and Right-Hand Side. Colours indicate crystal form shown in legend with blank squares having 0 crystals. Numbers correspond to the bottom panel crystal images.

From optimisation 2 the diffracting crystal was grown in the presence of pH 6 MES and 1500 mM ammonium sulphate. Further optimisation rounds were carried out on this condition to narrow down the crystallisation window for the two crystal forms. Both robotic and manual crystallisation trials were performed with optimal crystals from robotic screening. The two final conditions were 1550 mM ammonium sulphate, MES pH 5.8 for pyramids, and 1500 mM ammonium sulphate, MES pH 5.9 for plates, both with 2% v/v glycerol within the reservoir solution and 5% within the protein solution (figure 3.2.4-3). Crystals from the two conditions were mounted onto cryo loops and exposed to synchrotron radiation at Diamond Light Source with the outermost Bragg spots at 16.5 Å.

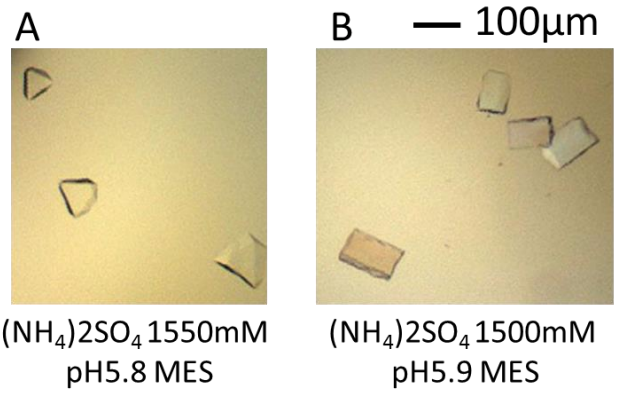


Figure 3.2.4-3**: Optimal pyramid (A) and plate (B) shaped crystal forming conditions for NTD_{31-145}Strep.** Crystals were grown at 19°C in the presence of 2% v/v glycerol within the reservoir and 5% v/v glycerol added to the protein solution prior to plate dispensing.

3.2.4.1 Silver Bullets additive screen

X-ray diffraction had so far been optimised by screening around the original crystallisation condition improving by more than 2-fold on the “plate” and “pyramid” crystals harvested. Additive screens are a portfolio of small molecules known for their ability to improve crystal formation, packing, and change crystal form (Culurgioni et al., 2010). The Silver Bullets additive screen contains a mixture of between 1 and 20 similar compounds in a single condition allowing testing of a cocktail of additives in a single crystal drop. Similarities between the 96 cocktails narrow down the compounds which help crystal formation with general trends between conditions.

To test the effects of the additives on both the “plate” and “pyramid” crystals, their buffer conditions were mixed and dispensed manually into four 96 well plates. Using the Mosquito® multi-aspirate technique 400 nL of protein was dispensed with 25 nL of additive + 175 nL of reservoir for each for the two crystal forms. Control plates were also dispensed with all plates sealed, stored at 19°C and imaged daily.

Comparison between the control and the silver bullet plates showed many additives had little effect on crystals. Some additives swapped the crystal form from “pyramid” to “plate” with others encouraging crystal formation where crystals were absent in the control. In one condition the additive

changed crystal form from pyramids to plates with one further cylinder crystal form present (figure 3.2.4.1A). Crystals were mounted onto cryo loops covering a wide selection of conditions with cryo-protection for X-ray data collection at Diamond Light Source I24. Additives did not improve any of the pyramid crystals that were screened. The "plate" crystals did improve with additives known for dehydrating the crystal with a new diffraction pattern seen with a repeating 4:2 pattern of Bragg maxima (figure 3.2.4.1C). Further improvements were seen for the new "cylinder" crystal form with a greater distribution and amount of diffraction maxima over a single image and over a data collection (figure 3.2.4.1B). For this cylinder crystal (figure 3.2.4.1A) data were collected and auto processed to 11.5 Å, the presence of Zinc was also detected by X-ray fluorescence scan with a peak at 1.2823 Å and inflection at 1.2827 Å.

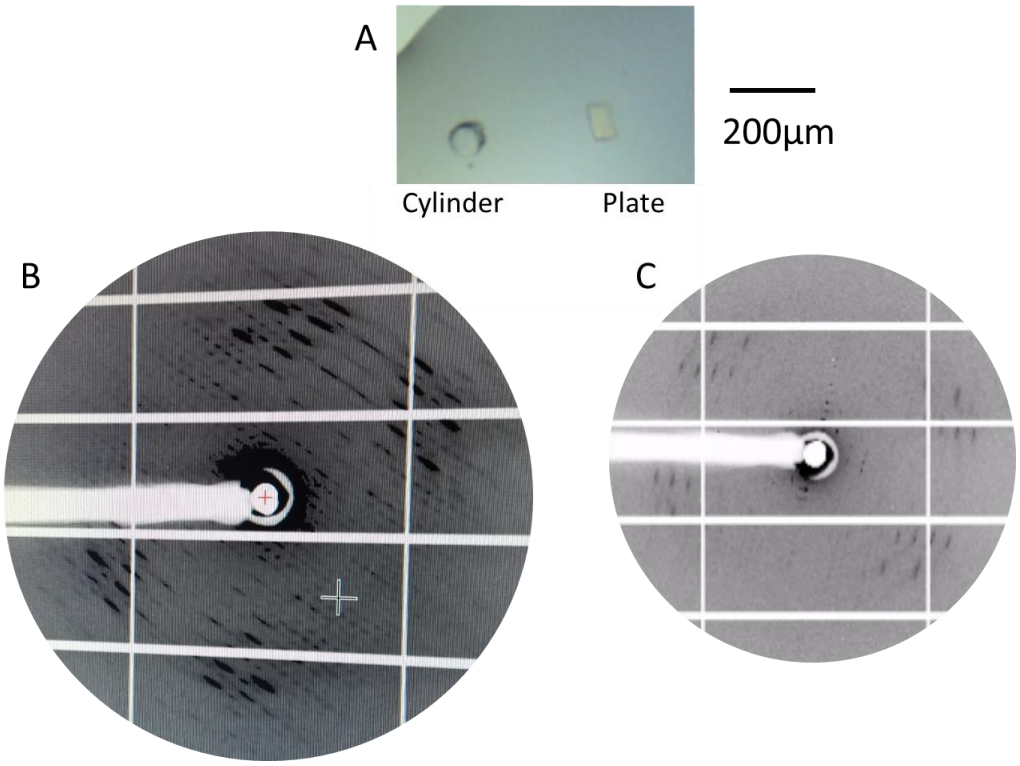


Figure 3.2.4.1: **NTD_{31-145}Strep Crystals show 4:2 Bragg diffraction pattern. A)** A new crystal form was seen with a cylindrical appearance alongside the previously seen 'plate' crystal form. **B)** Diffraction obtained from the 'cylinder crystal in A. The edge of this image is 11.5 Å. **C)** Diffraction obtained from the 'plate' crystal in A, the edge of this image is 13 Å.

3.2.4.2 I24 Dehydration HC1 & in-situ data collection

Further optimisation was necessary to improve crystal diffraction with the additive screen conditions alluding to dehydration improving the "plate" crystals. In-situ dehydration of crystals using the HC1 dehumidifier and direct X-ray data collection allows analysis of crystal diffraction with precise changes

in relative humidity. The diameter of the crystal is measured as an indication of hydration to determine and control of the rate and amount of crystal dehydration. The effect of dehydrating the crystal by X-ray diffraction can then be measured directly and experiment changed accordingly (Lobley et al., 2016). Unfortunately attempts to dehydrate crystals and measure X-ray diffraction were impeded by the low 11% humidity within the beam hutch. This resulted in a drastic reduction in X-ray diffraction during the process of crystal mounting, transfer and wicking. To ensure the starting crystals could diffract, in-situ plate crystal diffraction was collected. In contrast, diffraction was present to the same extent as previously seen for crystals collected at cryo conditions for both “plates” and “pyramids”. Other methods to dehydrate a crystal such as transferring the crystal into a higher percentage precipitant and additives were also tested, however diffraction did not improve.

3.2.4.3 Hampton additive screen

Crystallisation solutions were further optimised to give cylinder crystals without the need for additives (figure 3.2.4.3A). As further screening of Silver Bullets additives gave no additional improvement a change of approach was taken. The Hampton additive screen is composed of 96 individual conditions with the additive included in the reservoir condition enabling easier optimisation (Hampton Research, 2001).

The starting point for the Hampton Additive screen was 1450 mM ammonium sulphate, 6.5% v/v glycerol, 100 mM MES pH 5.9 with 5% v/v glycerol in the protein solution. This condition was made at 110% to allow 10% individual addition of each of the 96 additives. A second control plate was made with 10% water instead of additive. Both plates were dispensed and imaged using the standard protocol. From the additive screen maltose and sucrose gave more defined edges to “plate” crystals. These were further tested with no improvement over the previous 11.5 Å resolution. Few conditions gave rise to a swap in crystal form however the pyramid crystals still gave poor diffraction. One further crystal form appeared in the presence of sodium thiocyanate, this crystal grew from a small piece of dust into a 2D plate (figure 3.2.4.3B). The crystal was mounted onto a cryo loop, cryo-protected and exposed to X-rays at the Diamond Light Source synchrotron beamline IO3. The diffraction pattern showed similarities to the plate 4:2 Bragg diffraction, although with separated spots (figure 3.2.4.3C). With collection of a 1000 image dataset auto processing now resulted in a resolution of 9.77 Å.

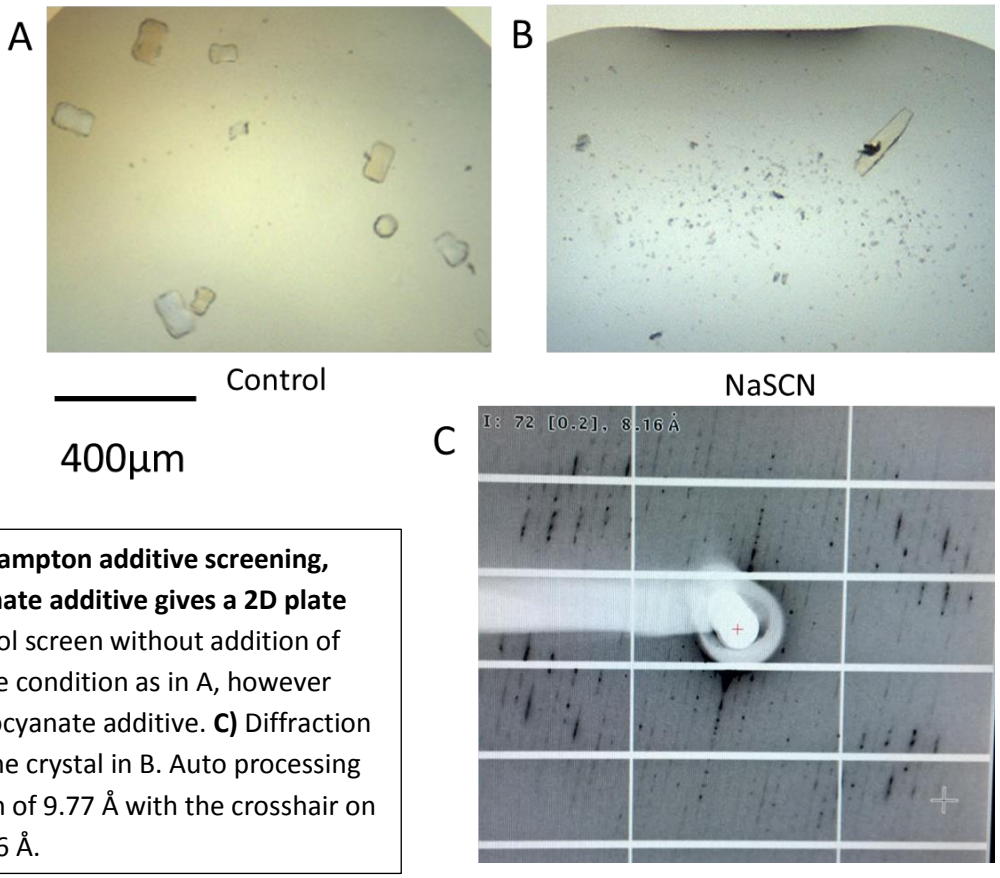


Figure 3.2.4.3: **Hampton additive screening, sodium thiocyanate additive gives a 2D plate crystal. A)** Control screen without addition of additive. **B)** Same condition as in A, however with sodium thiocyanate additive. **C)** Diffraction obtained from the crystal in B. Auto processing gave a resolution of 9.77 Å with the crosshair on the image at 8.16 Å.

3.2.4.4 Sodium thiocyanate crystal seeding methods

Further optimisation rounds were performed with small rod-shaped crystals appearing in low concentrations of sodium thiocyanate (figure 3.2.4.4A). Variation of the crystal condition failed to increase the crystal size. Since small crystals did appear these were used as starting seed material to encourage crystal growth. A seed stock was generated from the crystals in figure 3.2.4.4A as described in chapter 2.

Traditional streak seeding was used both with manual 24 well plates and Mosquito® dispensed 96 well plates with no crystalline material present after 31 days. Using the Mosquito® multi-aspirate seeding technique can be used with a dilution series of seeds from 10-, 100- and 1000-fold with 1:100 seed dilutions giving the best results. Rather than rod-like crystals within the solution, crystals were now forming on the surface of the plate with a wedge shape (figure 3.2.4.4B). Crystals were cryo-protected by quickly passing through a 20% v/v glycerol 80% v/v reservoir solution when mounted onto cryo loops prior to freezing. Data were collected at the Diamond Light Source synchrotron on beamline IO3 with a resolution improvement to 7.86 Å. The same diffraction pattern as the previous sodium thiocyanate crystal with spots overlapping the Bragg 4:2 diffraction was also maintained.

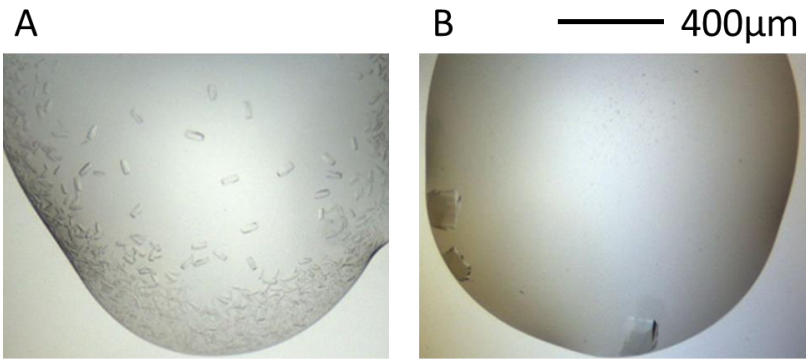


Figure 3.2.4.4: **Mosquito® seeding for sodium thiocyanate crystals from Hampton additive screen. A)** Crystals used for the seed stock. **B)** Crystals obtained from Mosquito® seeding.

3.2.4.5 Variation of temperature

Temperature can be a significant variable in crystallisation due to its ability to change solubility, supersaturation and speed of crystal growth (Hampton Research, 2001). So far, all crystallisation trials were carried out at 19°C. Therefore 19°C, 11°C and 4°C were tested using identically dispensed plates. Optimisation screen 18 was used with sodium thiocyanate and a 1:100 seed stock (appendix figure 7.3). The “cylinder” crystal form gave the best improvements, with visual changes from a cylinder to a hexagonal prism with sharp defined edges. This is apparent in figure 3.2.4.5 comparing from left to right as the temperature reduces, UV detection confirmed that these new hexagonal prism crystals contained $NTD_{31\text{-}145}$Strep.

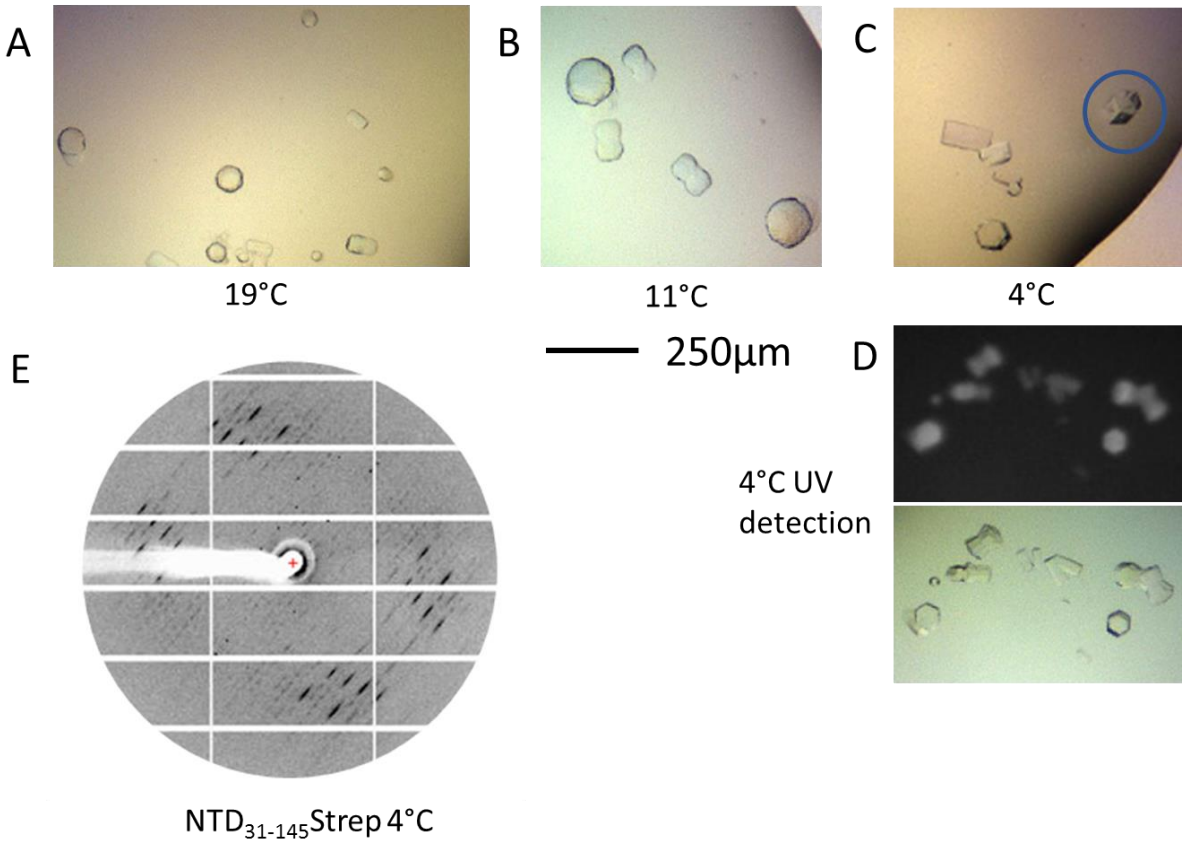


Figure 3.2.4.5**: Temperature variation improving cylinder crystal to a hexagonal prism.** For the three identical plates **A)** 19°C**, B)** 11°C, **C)** 4°C. Crystal circled was mounted for X-ray diffraction in E. **D)** UV detection of the 4°C crystal shows they contain protein. **E)** X-ray diffraction pattern of the hexagonal prism circled in blue, outer edge of diffraction in this image is 8.0 Å.

Hexagonal prism crystals were mounted onto cryo loops and cryo-protected by quickly transferring into a 20% v/v glycerol 80% v/v reservoir solution prior to flash freezing. Crystals were analysed by X-ray diffraction at the Diamond Light Source Synchrotron on beamline IO3. Figure 3.2.4.5E shows the diffraction obtained from the crystal circled in blue with AutoProc processing giving an outer shell resolution of 8.45 Å. Further X-ray diffraction maxima were seen at 4.63 Å in a subset of the images however there were no signs of diffraction in-between. A zinc fluorescence scan gave a peak at 1.2816 Å showing the presence of zinc within the crystal with further resolution optimisation necessary.

3.2.4.6 Purification of Selenomet NTD_{31-145}DM-Strep

Generation of phases from the zinc ions present within the crystals was limited by the diffraction resolution. Expression of protein containing heavier atoms such as selenium is one method to

generate phases. Generally, selenomethionine (SeMet) incorporation leads to a lesser diffracting crystal. Nevertheless, a change to the protein and its surroundings has the potential to improve diffraction. $NTD_{31\text{-}145}$Strep contains three methionine residues and so a SeMet labelled preparation was performed. The purification deviated from the non-labelled purification to allow addition of 3 mM TCEP with an amylose initial capture. Figure 3.2.4.6 shows the final polishing step of the purification with fractions containing $NTD_{31\text{-}145}$Strep pooled. For final SEC both with and without 2 mM DTT was performed with both samples dispensed into crystallisation trials.

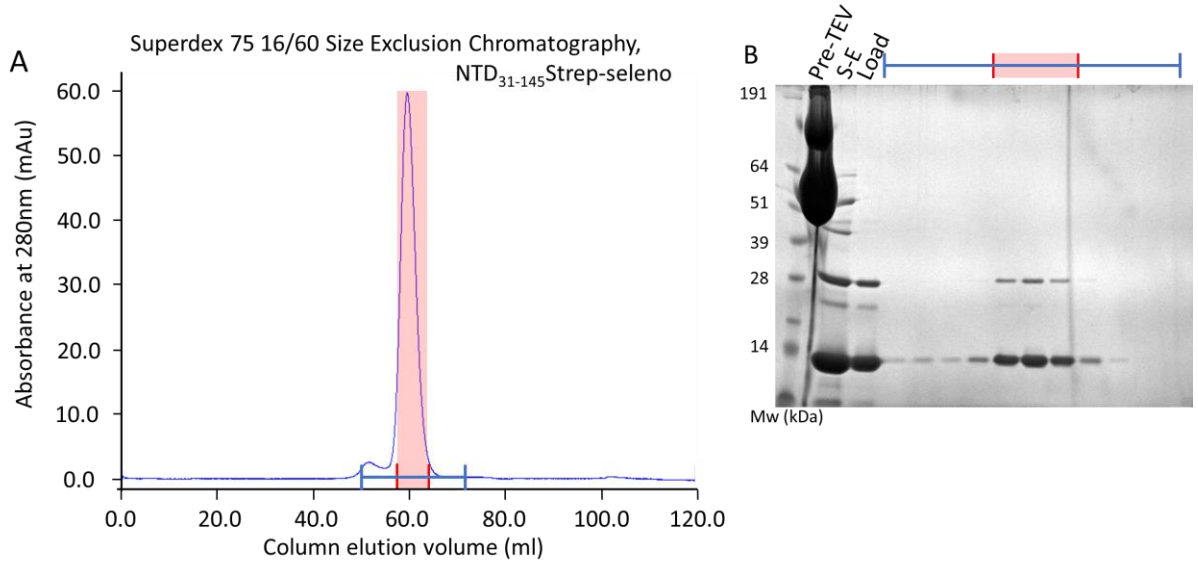


Figure 3.2.4.6: **SEC purification of SeMet $NTD_{31\text{-}145}$Strep.** Protein was purified using the standard procedure with adaptions detailed in the main text. **A)** SEC chromatogram of SeMet $NTD_{31\text{-}145}$Strep using the Superdex 75 16/60 SEC column. Fractions were analysed by SDS-PAGE in B. **B)** Protein eluted from the initial amylose column (pre-TEV) was further processed using the standard procedure with Streptactin elution (S-E) and protein applied to the SEC column (Load) shown. Fractions from SEC are shown with fractions in red pooled and concentrated to 10 mg/mL for crystallisation trials.

To encourage crystal formation non-labelled micro seeds were used in half of the crystallisation trials, with Wizard 1&2, Morpheus, JCSG and optimisation 1/18 (appendix 7.2/7.3) screens dispensed for both with and without DTT supplemented protein. Crystal plates were incubated at 4°C to give the best chance of crystals with identical screens at 19°C. Over all the different crystallisation conditions there were zero signs of crystal formation with absence of all previous crystal forms.

3.2.5 Surface entropy minimisation of NTD_{31-145}Strep

Further methods of optimisation were required to increase the diffraction. It is well known that protein crystallisation is dependent on formation of an ordered lattice. This is countered by the entropic cost of ordering and the loss of side change conformational freedom involved in crystal contacts. Clusters of highly entropic amino acids such as lysine, glutamine and glutamic acid can therefore be detrimental to this crystallisation process. A strategy to improve and encourage protein crystallisation is to therefore reduce patches of the surface entropy by mutational surface engineering. This should then lower the energy barrier for crystal formation (Derewenda and Vekilov, 2006). Examples from the literature show surface entropy minimisation providing an improvement in diffraction and crystal formation (Cooper et al., 2007; Derewenda and Vekilov, 2006; Parthasarathy et al., 2008; Rzechorzek et al., 2014).

3.2.5.1 Construct design

To determine key clusters of residues for surface entropy minimisation the online UCLA SER_p server was used in October 2017 (Goldschmidt et al., 2007). Here three patches were chosen for surface entropy minimisation named cluster 1 (C1), 2 (C2) and 3 (C3) each in a different location on the protein (figure 3.2.5.1). Mutations were mapped onto the known structures of Rad50 and Cortexillin with side chains visibly protruding from the alpha helix.

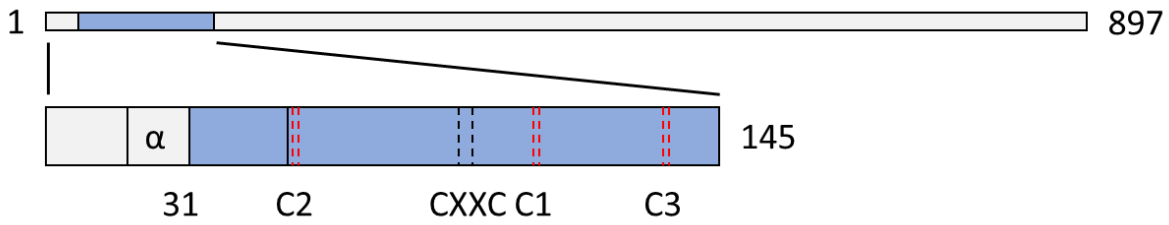


Figure 3.2.5.1: **Schematic drawing of NTD_{31-145}Strep cluster mutations.** NTD_{31-145}Strep is shown in blue, with tetramerisation domain (α) and zinc binding domain shown (CXXC). NTD_{31-145}C2-Strep (C2 = K46A, Q47A, E48A). NTD_{31-145}C1-Strep (C1 = K107A, K108A, Q109A). NTD_{31-145}C3-Strep (C3 = E135A, Q136A, K138A).

3.2.5.2 Recombinant protein expression and purification

Protein expression was performed in Rosetta™ 2 (DE3) with two litres of culture used per protein per purification. The same purification strategy as wild type protein was performed with initial Ni-NTA followed by overnight cleavage and secondary strep capture. NTD_{31-145}C2-Strep purification is shown in figure 3.2.5.2. SEC showed overlapping traces for the three clusters eluting at the same elution volume (figure 3.2.5.2E) indicating no gross change in shape from the mutations.

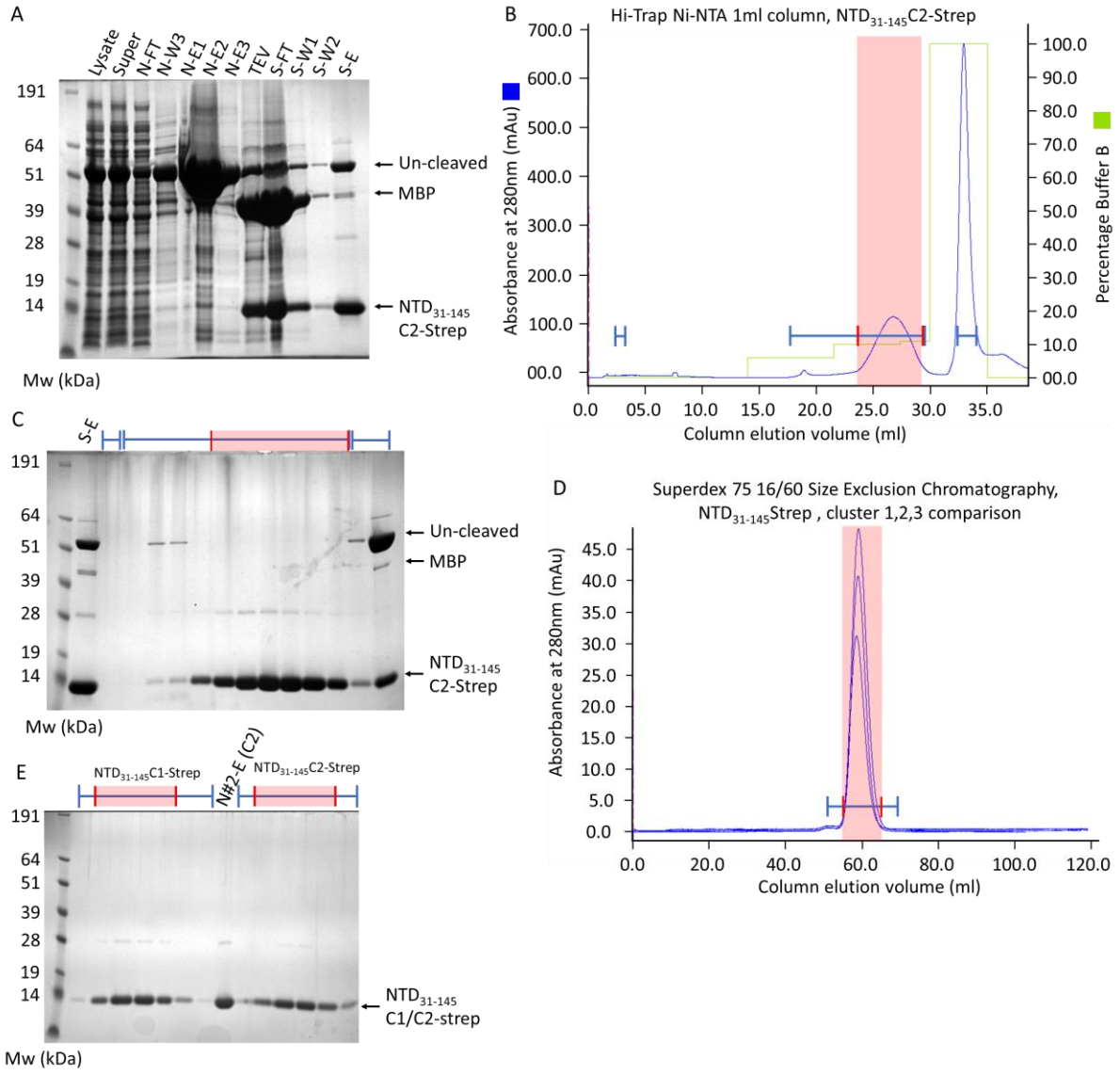


Figure 3.2.5.2**: Purification of NTD$_{31-145}$Strep Cluster mutants. A)** Initial purification of NTD$_{31-145}$C2-Strep through to elution from the Streptactin column. **B)** His-trap chromatogram. **C)** SDS-PAGE analysis of His-trap column with blue lines corresponding to fractions in B). Fractions highlighted in red were pooled for SEC in D/E. **D)** Superdex 75 16/60 SEC. Fractions for Cluster 1 and Cluster 2 were analysed by SDS-PAGE in E with traces overlaid for the three clusters. **E)** SDS-PAGE analysis of fractions of cluster 1 & 2 along with the protein loaded onto the column for cluster 2 (N#2-E (C2)). Fractions in red were pooled individually for C1 & C2 for concentration for experimental analysis.

3.2.5.3 Initial crystal screening

Crystal screens were dispensed for each NTD$_{31-145}$Strep cluster mutated protein, including Wizard™ 1&2, Morpheus® and JCSG+™ both with and without micro seeds. Zero crystals were seen without micro-seeds however micro seeding gave UV absorbing crystals in five conditions. A total of 32 crystals

were mounted and cryo-protected for data collection at Diamond Light Source, however there were no signs of protein diffraction. Many crystals had salt diffraction patterns, therefore UV absorbance was likely from protein deposited on the salt crystal surface. Thermal melt analysis was performed of the cluster mutated proteins showing a reduction in the melt peak compared to wild type protein. This indicated reduced stability and is likely the reason why the mutations failed to help with crystallisation (figure 3.2.5.3).

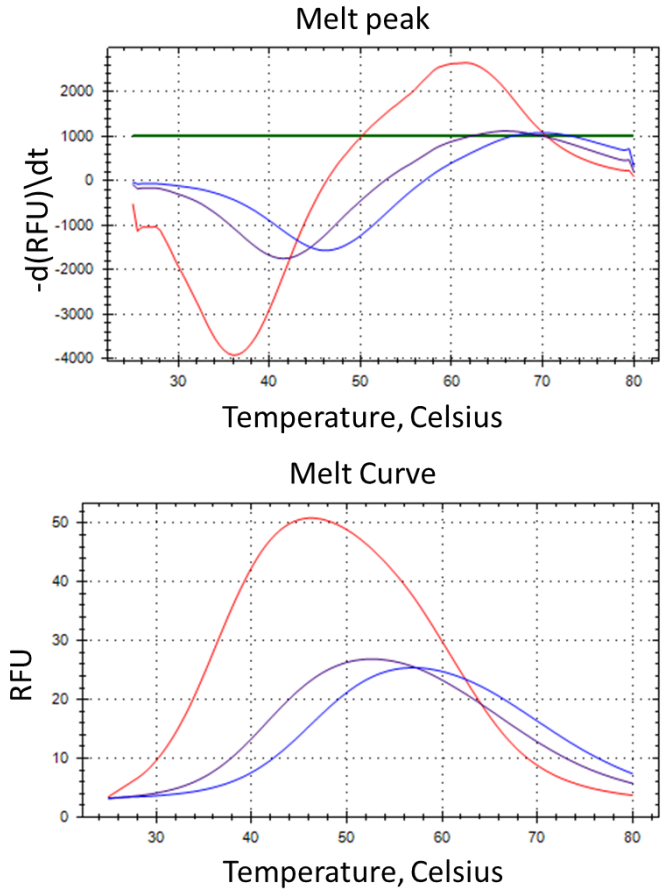


Figure 3.2.5.3: **Thermal melt peak and melt curve.** From the binding of SYPRO orange to unfolded protein the fluorescence increases and is measured in relative fluorescence units. Three repeats for each protein gave the same melt temperatures of 36°C (C2) 41.5°C (C1) and 46.0°C (Wild type)

3.2.6 Crystallisation of NTD_{31-145}DM-strep

The NTD_{31-145} is predicted to contain two coiled-coil regions separated by a zinc-binding motif based on amino acid sequence analysis. Previous studies of a double cysteine-to-alanine mutation targeting the putative zinc ligands C89 and C92 showed no change to the dimerization state of the CtIP-NTD (Davies et al., 2015). As incomplete or diverse incorporation of zinc by recombinant CtIP-NTD was a possible source of sample heterogeneity hampering crystallisation, a double mutant NTD_{31-145}DM-Strep construct replacing C89 and C92 with alanine was generated (figure 3.2.6).

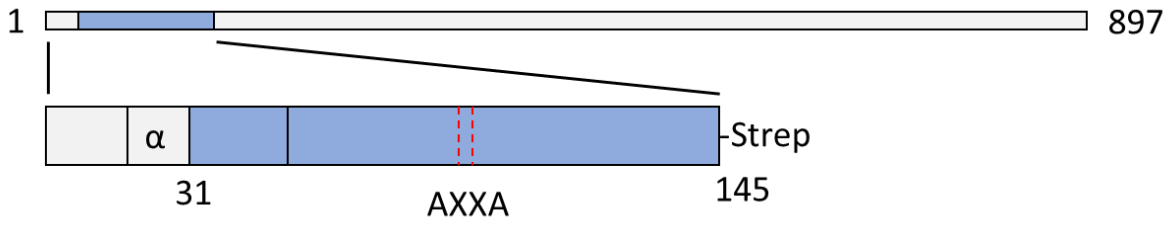


Figure 3.2.6: **Schematic drawing of NTD_{31-145}DM-Strep.** NTD_{31-}145DM-Strep is shown in blue, with tetramerisation domain (α) and zinc binding domain mutated to alanine's (AXXA). DM stands for C89A/C92A (AXXA) – Double Mutant.

3.2.6.1 NTD31-145DM-Strep protein purification.

Protein expression was performed in Rosetta™ 2 (DE3) cells transformed with NTD_{31-145}DM-Strep. Two litres of cells were used for each purification using the same protocol as for the wild type construct. NTD_{31-145}DM-Strep eluted from the His-Trap Ni-NTA column in the same concentration of imidazole although with a wider peak (figure 3.2.6.1). In the Superdex 75 16/60 SEC, NTD_{31-145}DM-Strep eluted in the same elution volume as the wild type protein (figure 3.2.6.1D), confirming previous results that the protein can still form a dimer without the incorporation of zinc. Eluted fractions from SEC containing NTD_{31-145}DM-Strep, as shown by SDS-PAGE, were pooled and concentrated to 200 µL. The Strep tag enabled a protein concentration of 9.8 mg/mL to be determined by absorbance at 280 nm. Protein was snap frozen with liquid nitrogen in 20 µL aliquots. No higher molecular weight dimer band was seen after SEC, with minimal appearance on the Hi-Trap Ni-NTA gel.

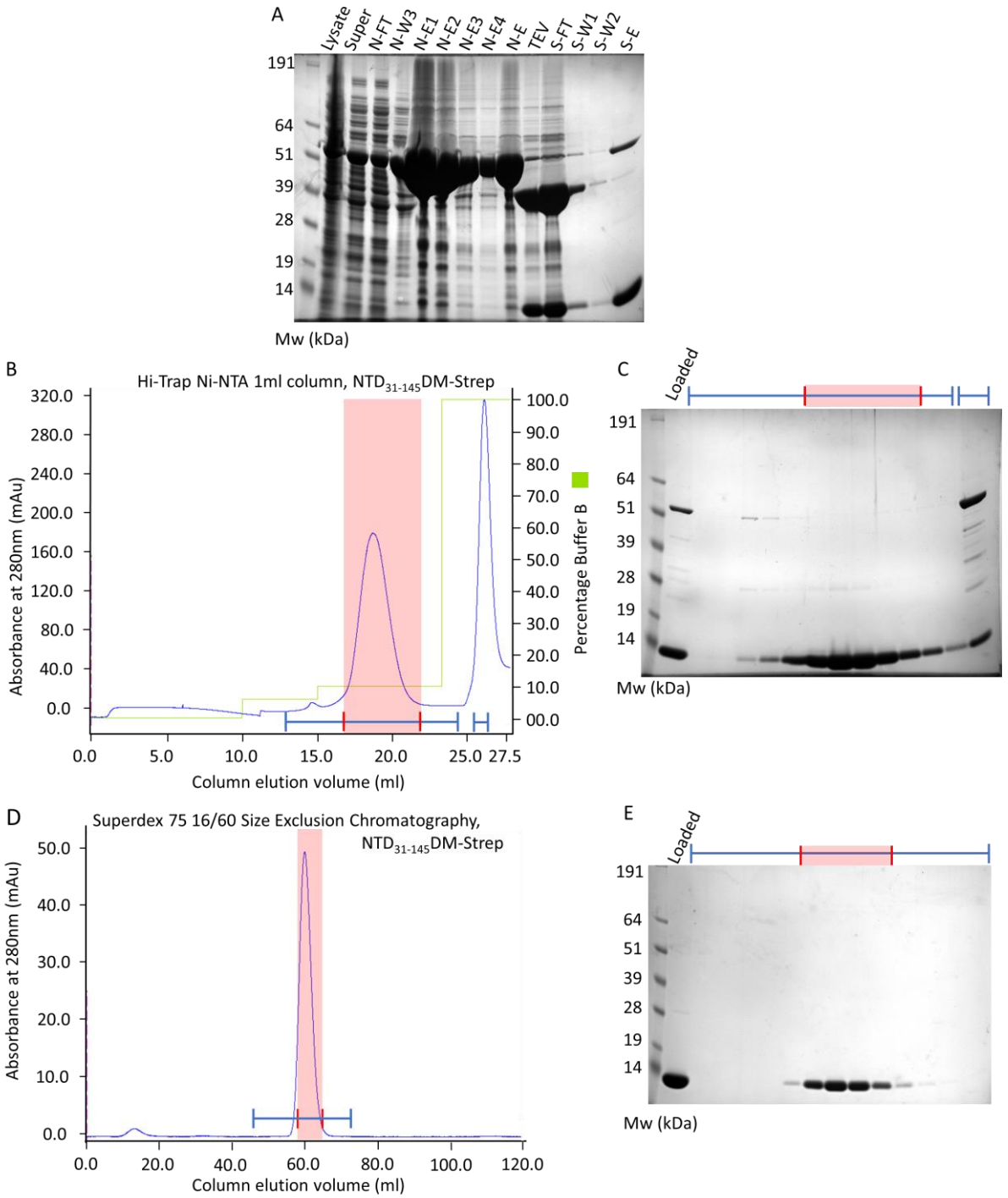


Figure 3.2.6.1: **Protein purification of NTD_{31-145}DM-Strep**. **A)** SDS-PAGE analysis of initial capture with Ni-NTA combined elution (N-E) pooled for subsequent TEV cleavage and Strep-Tactin capture. Strep combined elution (S-E) was loaded into the Hi-Trap Ni-NTA column in B. **B)** Chromatogram for the His-Trap Ni-NTA, intrinsic NTD_{31-145}DM-Strep affinity to the column allowed purification eluting with 10% v/v buffer B (1M NaCl) with fractions analysed by SDS-PAGE in C. **C)** Fractions highlighted in red were pooled for SEC in D. **D)** Superdex 200 16/60 SEC of pooled Ni-NTA elution from C with SDS-PAGE analysis in E. **E)** fractions highlighted in red were pooled and concentrated to 10 mg/mL prior to freezing for future experimentation.

3.2.6.2 Crystal screening of NTD_{31-145}DM-Strep

The Morpheus® (appendix 7.8) and Wizard™ 1&2 crystallisation screens were dispensed using a 1:1 and 2:1 protein to reservoir drop ratio at a protein concentration of 7.5 mg/mL. Crystallisation plates were stored at 19°C and imaged daily. In contrast to the wild type protein, there were no hits for the Wizard 1&2 screen. For the Morpheus® screen, only condition F6 (appendix 7.8) yielded a crystal, which continued to grow for 8 days (figure 3.2.6.2-1). At this point it was mounted onto a cryo loop without further cryo-protectant from the 400 nL 1:1 ratio drop. The crystal was very fragile and broke into two parts.

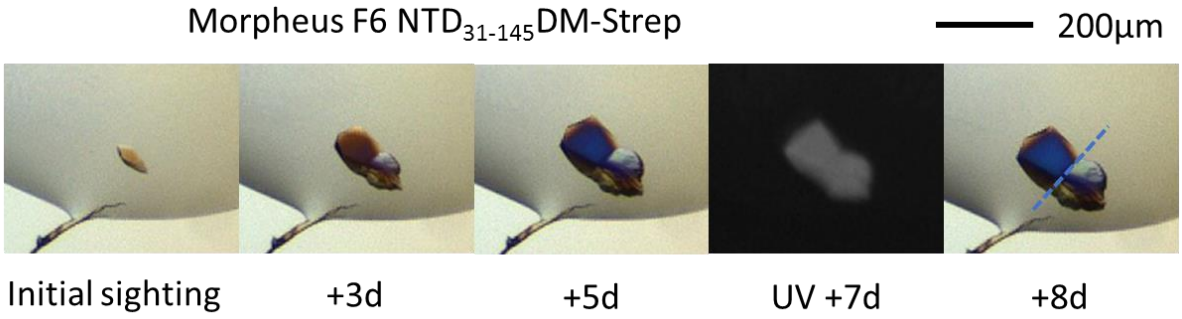


Figure 3.2.6.2-1**: Growth NTD_{31-145}DM-Strep crystal F6 Morpheus®.** See Appendix 7.8 for crystal conditions. The blue dashed line indicates where crystal broke. Fast UV was performed to detect UV absorbance of the crystal. Images are from first detection at Day 0 + 12 hours through to +8

The two mounted parts of the crystal were screened at beamline IO4 of Diamond Light Source. The NTD_{31-145}DM-Strep crystal showed strong diffraction to almost 3A resolution (figure 3.2.6.2-2a/b), confirming that this crystal form was suitable for single-crystal X-ray diffraction analysis.

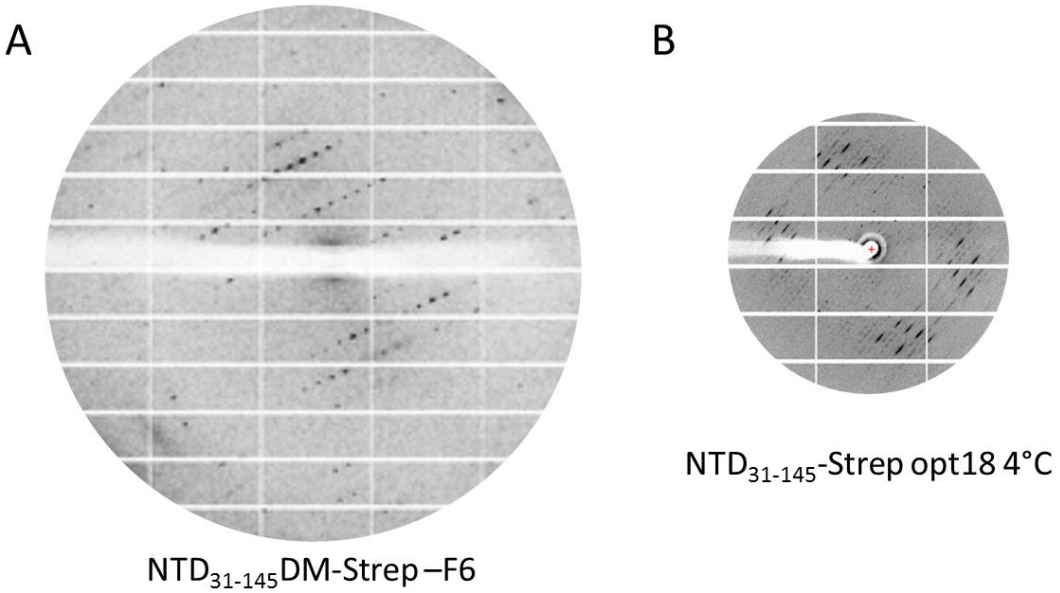


Figure 3.2.6.2-2**: X-ray Diffraction of NTD_{31-145}DM-Strep compared to NTD_{31-145}-Strep**. **A)** Diffraction pattern of NTD_{31-145}DM-Strep crystals formed in well F6 of Morpheus screen, the outer edge of the diffraction in this image is 3.5 Å. Auto processing gave a space group of $P3_1$ with A 86.81, B 86.81, C 42.89, α 90.00, β 90.00, γ 120.00 unit cell dimension. **B)** Diffraction pattern of NTD_{31-145}Strep hexagonal crystal formed in optimisation 18 at 4°C, outer edge of diffraction in this image is 8.0 Å.

3.2.6.3 NTD_{31-145}DM-Strep Matrix Micro Seeding

Upon repeat of the Morpheus® crystallisation condition, no new crystals were obtained. To promote crystal formation a seed stock was made from the crystal that had been exposed to synchrotron radiation (methods 2.19.1.4). The technique of Matrix Micro Seeding was used to directly seed into the Morpheus® crystallisation screen using the multi-aspirate seeding technique (D'Arcy et al., 2007). Two crystallisation trials were performed both with and without seed stock, a constant 1:1 protein to crystallisation condition drop ratio and a 400 nL drop volume.

For the crystallisation trial without seeds, crystals were seen in wells D7, E11 and F11 (figure 3.2.6.3-1). The crystals from D7 and F11 broke into two parts giving two attempts at data collection for each crystal. The change of additive from monosaccharides in F11 to ethylene glycols made crystal E11 less fragile and therefore did not break (see appendix 7.8 for conditions).

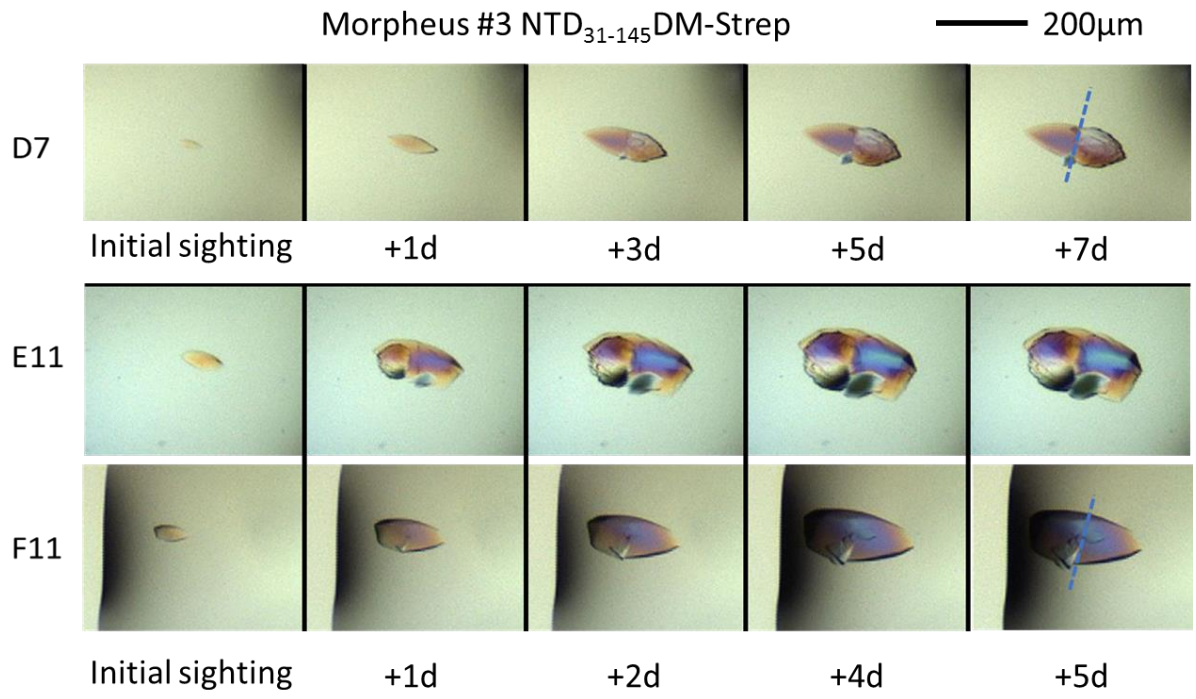


Figure 3.2.6.3-1**: Growth of NTD_{31-145}DM-Strep crystals formed in Morpheus repeat #3.** Crystals formed in condition D7, E11 and F11 (see Appendix 7.8 for crystal conditions). The blue dashed line indicates where crystals were broken to enable two data collections. Crystals were initially seen at +12 hours. The crystal from condition D7 was harvested after 7 days (d), whereas the crystals in conditions E11 and F11 were harvested after 5 days.

From matrix micro seeding 65 out of 96 conditions yielded crystalline material; half of these gave crystals of sufficient size to be tested for X-ray diffraction (figure 3.2.6.3-2). Like the wild type protein, NTD_{31-145}DM-Strep tended to crystallise under basic pH conditions. 32 Crystals were of sufficient size to be harvested after 4 days. Morpheus® conditions already contain additives to provide cryo-protection so no further cryo-protection was added on crystal mounting. Most of these crystals were attached to the plate surface and were very delicate, making crystal mounting difficult, especially in conditions containing alcohols.

Upon X-ray diffraction screening of the non-seeded crystals, the cryo loops contained excessive ice, using grid screening over the cryoloop surface, diffraction could be seen through the ice indicating the presence of protein crystal diffraction the same as previous data collections.

Figure 3.2.6.3-2**: Crystals formed in Morpheus® screen with Matrix Micro Seeding.** The crystal plate was stored at 19°C and imaged daily by RockMaker. See appendix 7.8 for crystal conditions. Crystal diameter along the longest axis was measured to classify, ≤40 µm were classed as "tiny Crystals", ≤60 µm were classed as "small crystals". Crystals that were ≥80 µm are depicted, with blue circles indicating a selection of mounted or important crystals.

3.2.6.4 SeMet NTD_{31-145}DM-Strep protein purification

As the zinc binding site had been mutated, zinc was no long present in the crystal and could not be used to obtain phase information using its anomalous scattering signal. Therefore, a SeMet-labelled protein preparation was performed.

Protein expression was performed using Rosetta™ 2 (DE3) bacterial cells transformed with NTD_{31-145}DM-Strep. Four litres of cells were grown in minimal media supplemented with SeMet as detailed in methods 2.1.4. The standard purification protocol was adapted to use a 5 mL amylose column for initial capture enabling addition of reducing agent. TEV overnight cleavage and strep capture were performed as previously described. SeMet $NTD31_{-145}$DM-Strep eluted from the 1 mL His-Trap Ni-NTA column in the same concentration of imidazole with a similar profile to the non-labelled protein. Fractions corresponding to NTD_{31-145}DM-Strep were analysed by SDS-PAGE (figure 3.2.6.4C) and pooled for final polishing by Superdex 75 16/60 SEC (figure 3.2.6.4D/E). Eluted NTD_{31-145}DM-Strep was identified by SDS-PAGE, pooled and concentrated to 240 µL at 9.8 mg/mL for crystallisation experiments.

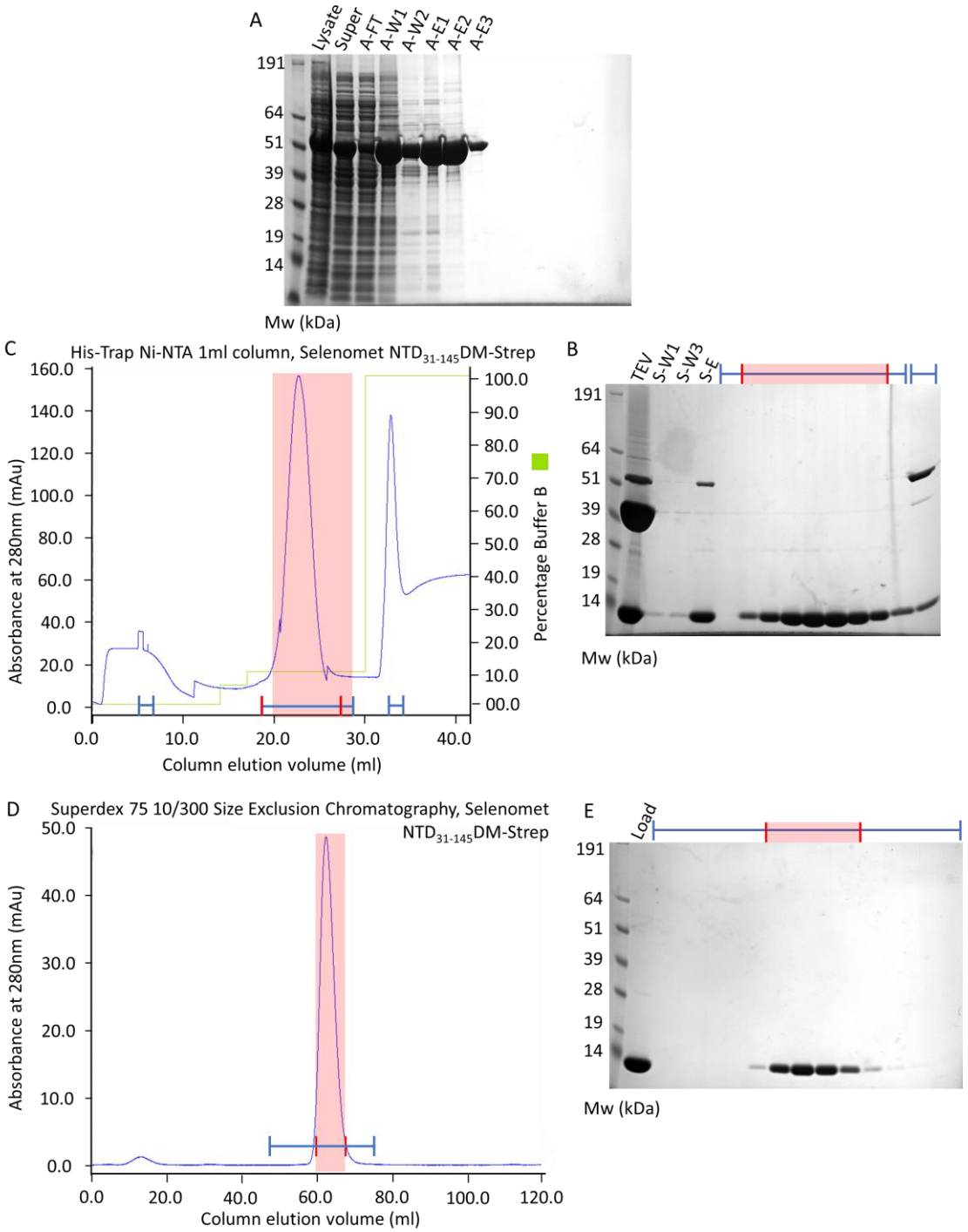


Figure 3.2.6.4: **Protein purification of SeMet NTD$_{31-145}$DM-Strep. A)** SDS-PAGE analysis of initial capture with Ni-NTA elution fractions pooled for subsequent TEV cleavage and Strep Tactin capture in B. **B)** Combined strep elution's (S-E) were applied to the Hi-Trap Ni-NTA column in C, with fractions indicated in B corresponding to C. **C)** Chromatogram for the Hi-Trap Ni-NTA, intrinsic SeMet NTD$_{31-145}$DM-Strep affinity to the column allowed purification eluting with 10% buffer B with fractions analysed by SDS-PAGE in B. Fractions highlighted in red were pooled for SEC in D. **D)** Superdex 200 16/60 SEC of pooled Ni-NTA elution from C with SDS-PAGE analysis in E. **E)** Fractions highlighted in red were pooled and concentrated to 10 mg/mL prior to freezing for experimentation.

3.2.6.4.1 SeMet NTD31-145DM-Strep Matrix micro seeding

SeMet NTD_{31-145}DM-Strep was dispensed into two Morpheus screens, with and without micro seeds, at a constant concentration of 7.5 mg/mL with 5% v/v glycerol. For both screens the drop ratio was the same as previous crystallisation trials with plates were stored at 19°C and imaged every 12 h for 10 days. In the control non-seeded plate only two amorphous spherulite crystals were present. Whereas in the seeded screen, addition of protein crystal nuclei resulted in 86 out of 96 conditions containing crystals. General crystallisation trends were seen similar to wild type protein; figure 3.2.6.4.1 shows examples of crystals formed.

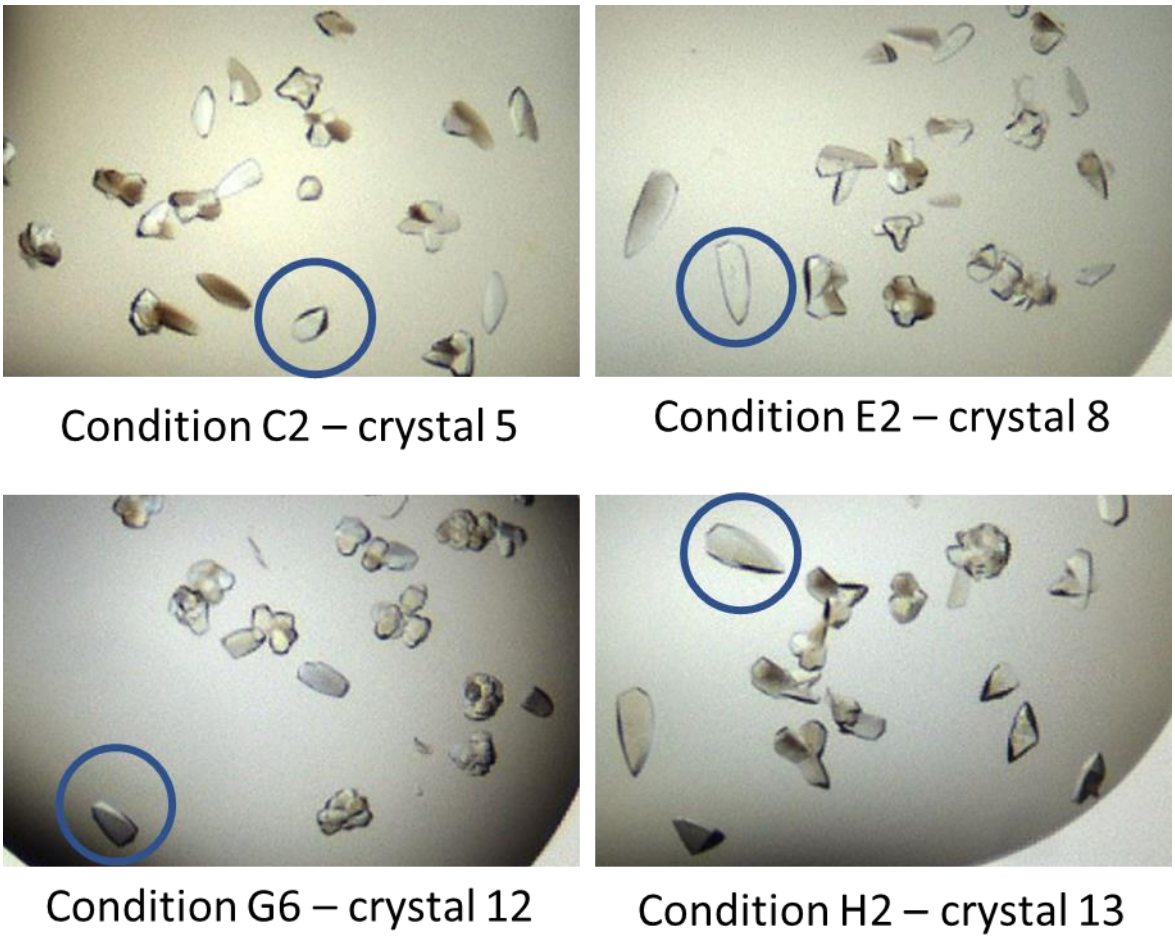


Figure 3.2.6.4.1**: SeMet NTD_{31-145}DM-Strep crystals.** Circle represents a 200 µm diameter and indicates crystals mounted for X-ray diffraction.

A total of 32 SeMet NTD_{31-145}DM-Strep crystals were screened for X-ray diffraction at beamline IO3 of Diamond Light Source. Many of the crystal loops contained excessive ice with grid screening over the cryo-loop obtaining diffraction to around 9.0 Å (E2 and H2 figure 3.2.6.4.1). Two crystals (C2 and G6 figure 3.2.6.4.1) diffracted to ≤5.0 Å however they suffered greatly from radiation damage with insufficient diffraction for generation of phase information.

3.2.7 Molecular Replacement

Crystals from the Morpheus seeded plate (section 3.2.6.3) were sent to the ESRF for X-ray data collection. X-ray diffraction data were collected at beam wavelength 0.976251 Å and 100K at 69.28% transmission as 990 contiguous frames with 0.02 second exposure and 0.1° oscillation on a Pilatus3_6M detector at beamline ID30B of the ESRF Grenoble, France. The diffraction images were indexed and intensities extracted and scaled in XDS (Kabsch, 2010) and further scaled and analysed using Aimless (Evans and Murshudov, 2013) with help from Prof. Luca Pellegrini . All crystals were screened, and 6 crystals provided complete datasets to 3.0 Å resolution or better. All data sets were in the same trigonal $P3_1$ space group with unit cell dimensions: a=b=86.6A c=42.6A, α=β=90, γ=120. Two copies of the $NTD_{31\text{-}145}$DM sequence forming one dimer were seen in the asymmetric unit. A crystal and an example diffraction image are shown in figure 3.2.7-1.

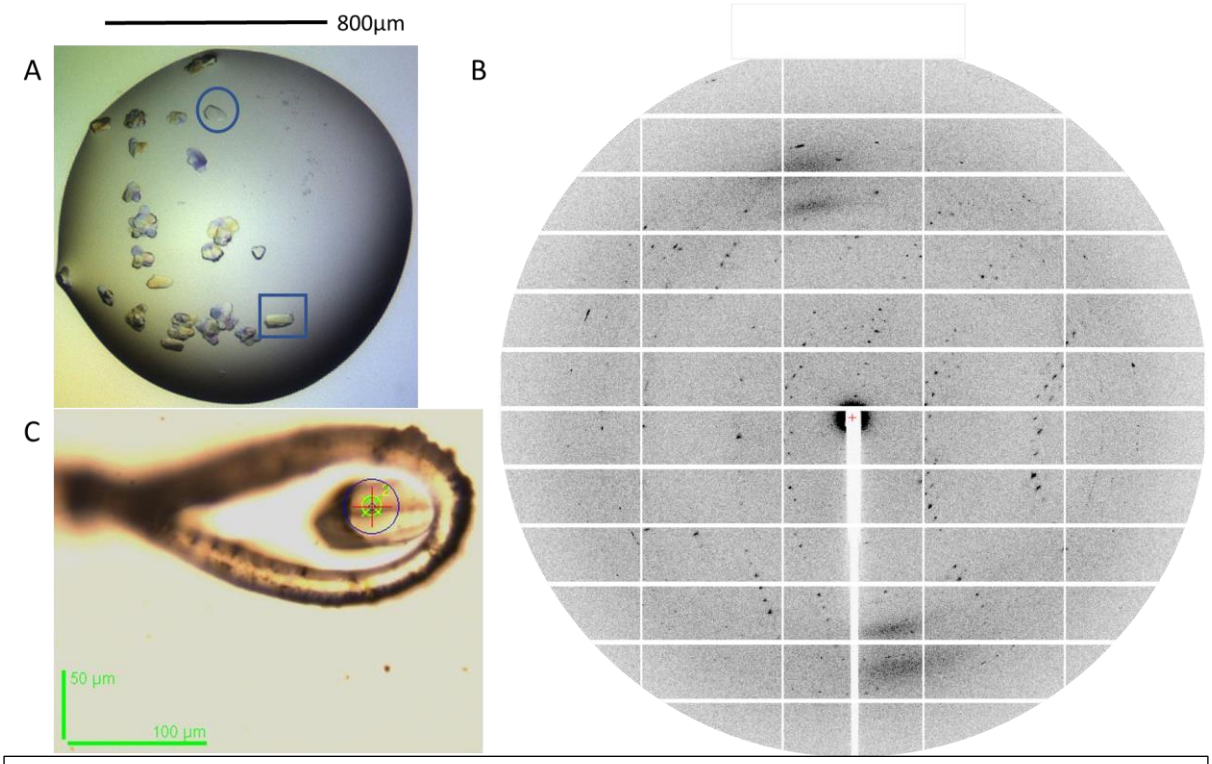


Figure 3.2.7-1: **Example crystal and diffraction pattern for $NTD_{31\text{-}145}$DM-Strep. A)** Crystal of $NTD_{31\text{-}145}$DM-Strep grown in the condition G6 Morpheus (molecular dimensions). Crystals indicated were mounted and exposed to X-rays. **B)** Crystal of $NTD_{31\text{-}145}$DM-Strep immobilised in the cryo-loop. **C)** Representative X-ray diffraction pattern of $NTD_{31\text{-}145}$DM-Strep, the resolution of the outer edge of this image is 2.58 Å.

The structure of NTD_{31-145}DM was determined by Molecular Replacement (MR) using Phaser (McCoy et al., 2007) as implemented within the CCP4 suite of crystallographic software (Collaborative Computational Project, 1994). Two helical coiled-coil models spanning amino acids 56-83 and 99-144 of human CtIP, corresponding to the regions of CtIP NTD strongly predicted to form coiled-coil structures, were generated in CBuilder2.0 (Wood and Woolfson, 2018) and used as search templates. Phaser successfully placed both models in the asymmetric unit, with no model clashes and clear indication of extra electron density for the missing parts of the crystallographic model. An initial poly-alanine model was refined in Phenix (Adams et al., 2010), adding side chains and extending the model to boundaries of the interpretable density maps. Refinement in Phenix was interspersed with manual rebuilding in Coot (Emsley and Cowtan, 2004), to improve fitting to the electron density map and the stereochemistry of the model. The refined crystallographic model of the human CtIP NTD comprise amino acids 31 to 136 for both chains, with R_{work}/R_{free} = 0.2229/0.262, no Ramachandran outliers and a Clashscore of 2.45 (Davis et al., 2007) (see Table 3.2.7-2). Figure 3.2.7-3 shows an overlay of the 2Fo-Fc electron density map contoured at 1.4 RMSD with the final crystallographic model of the NTD_{31-145}DM.

Unit cell: 86.60Å 86.60Å 42.60Å 90.00° 90.00° 120.00°
Space group: $P3_1$

	Overall	InnerShell	OuterShell
Low resolution limit (Å)	43.30	43.30	2.95
High resolution limit (Å)	2.80	8.86	2.80
R_{merge}	0.065	0.033	0.718
R_{meas}	0.081	0.041	0.886
R_{pim}	0.047	0.024	0.513
Total number of observations	24003	700	3599
Total number unique	8686	273	1283
Mean((I)/sd(I))	9.5	25.1	1.5
Mn(I) half-set correlation CC(1/2)	0.997	0.998	0.575
Completeness	98.7	98.0	99.2
Multiplicity	2.8	2.6	2.8
Mean(Chi^2)	0.98	1.11	0.91

Average mosaicity: 0.15

$$Rmerge = \frac{\sum_{hkl}\sum_{i=1}^{n}|I_{hkl,i}-\bar{I}_{hkl}|}{\sum_{hkl}\sum_{i=1}^{n}I_{hkl}} \;;\; Rmeas = \frac{\sum_{hkl}\sqrt{\frac{n}{n-1}}\sum_{i=1}^{n}|I_{hkl,i}-\bar{I}_{hkl}|}{\sum_{hkl}\sum_{i=1}^{n}I_{hkl}} \;;\; Rpim = \frac{\sum_{hkl}\sqrt{\frac{1}{n-1}}\sum_{i=1}^{n}|I_{hkl,i}-\bar{I}_{hkl}|}{\sum_{hkl}\sum_{i=1}^{n}I_{hkl}}$$

REFINEMENT STATISTICS

Resolution range	37.5 - 2.802 (2.902 - 2.802)
Unique reflections	8676 (878)
Reflections used in refinement	8655 (877)
Reflections used for R-free	387 (38)
R-factor	0.2229 (0.3215)
R-free	0.2620 (0.3143)
Number of non-hydrogen atoms	1831
macromolecules	1809
solvent	22
Protein residues	214
RMS(bonds)	0.001
RMS(angles)	0.23
Ramachandran favored (%)	100.00
Ramachandran allowed (%)	0.00
Ramachandran outliers (%)	0.00
Rotamer outliers (%)	1.00
Clashscore	2.45
Average B-factor	99.69
macromolecules	99.97
solvent	77.10

Table 3.2.7-2: **X-ray data collection and crystallographic refinement of CtIP $NTD_{31\text{-}145}$DM**. R_{merge} equation is shown where I_{hkl} is the average of the symmetry related observations of a unique reflection. R_{meas} is the redundancy-independent merging R-value where n is the number of observations. R_{pim} is the precision-indicating merging R-value. R_{merge} in the top intensity bin (low-resolution bin) is an indicator of overall data quality. Mean ((I)/sd (I)) describes the ratio of signal/noise. $CC_{1/2}$ is the correlation coefficient between two data sets generated by randomly splitting the data into two halves and is a measure of internal consistent of the collected data. Completeness is the number of crystallographic reflections measured, expressed as a percentage of the total number. Multiplicity is the average number of times that each spot of the diffraction pattern is measured. R-factor = $\Sigma\,(|F_{obs} - F_{calc}|)\;/\;\Sigma|F_{obs}|$, where F_{obs} and F_{calc} are the observed and calculated structure factors. R-free = R-factor calculated using a randomly selected 5% subset of the diffraction data, omitted throughout the refinement. Root-mean-square (RMS) is the deviation of the final crystallographic model from expected values of bond lengths and angles.

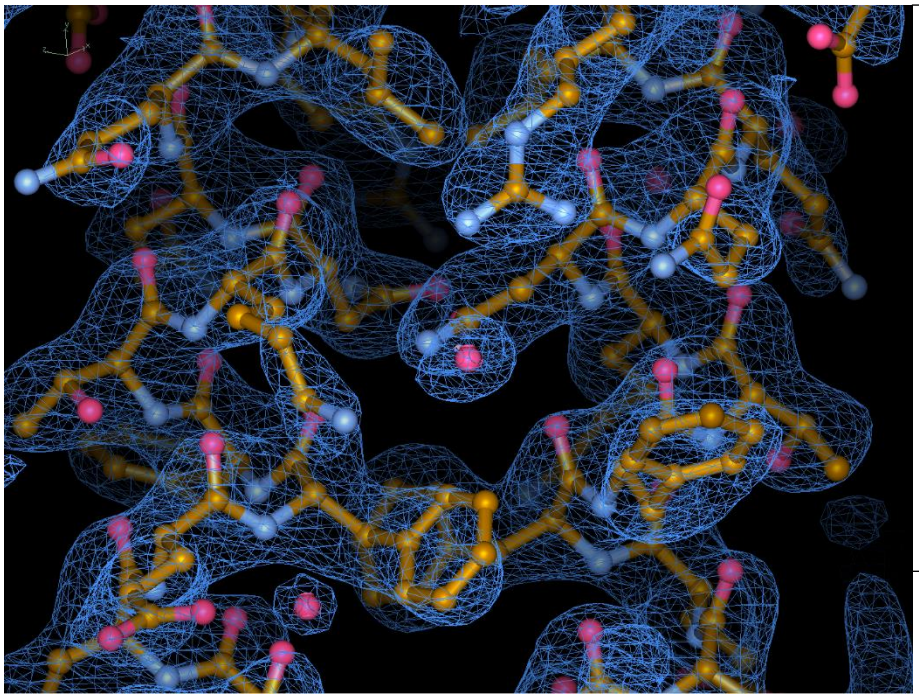


Figure 3.2.7-3: **Crystallographic model of $NTD_{31-145}DM$ overlaid with the 2Fo-Fc electron density map contoured at 1.4 RMSD** (root-mean-square deviation). Image was prepared using Coot (Emsley & Cowtan, 2004.)

The crystal structure revealed that the NTD region spanning residues 31-136 forms a highly elongated dimer of parallel uninterrupted alpha helices (figure 3.2.7-4). Within this helical structure, two coiled-coil regions (amino acids 31-85 and 110-136) form on either side of the zinc binding site; the residues surrounding the cysteine pairs, which act putatively as zinc ligands, forms instead two alpha helices that run alongside each other in the dimer. Formation of the first coiled coil agrees with a prevalence of hydrophobic residues at the '*a*' and '*d*' positions of the heptad repeat (figure 3.2.7-5), which disappears as the sequence progresses in the central parallel alpha helix region (figure 3.2.7-5A / figure 3.2.7-6A/C). Here (green figure 3.2.7-6A), the most prominent connection between the two alpha helices is a pair of phenylalanine residues (amino acid 106); their side chains form a π-stack that likely helps to form the start of the C-terminal coiled coil region (figure 3.2.7-6A).

Inspection of the structure shows that the position of the putative zinc-binding cysteines 89 and 92 falls within the central region of parallel alpha helices. Surprisingly, the alanine residues point towards the solvent and away from the inside of the coil, not as would be expected for shared coordination of a zinc ion. The reason for this unexpected conformation is currently unknown. It is possible that in the wild-type sequence the cysteines are part of local random-coil motif that lets them coordinate the zinc, and that in the mutant protein the presence of alpha-promoting alanine residues has caused formation of a continuous helical region (figure 3.2.7-6B/D). Further experiments will be necessary to determine the structure of the zinc-binding motif of the CtIP-NTD.

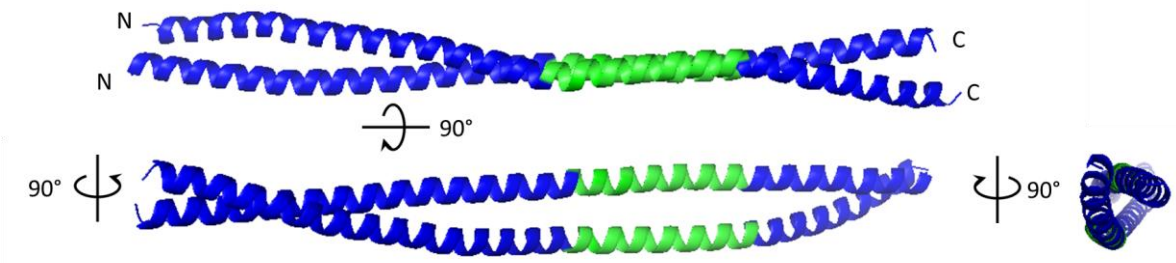
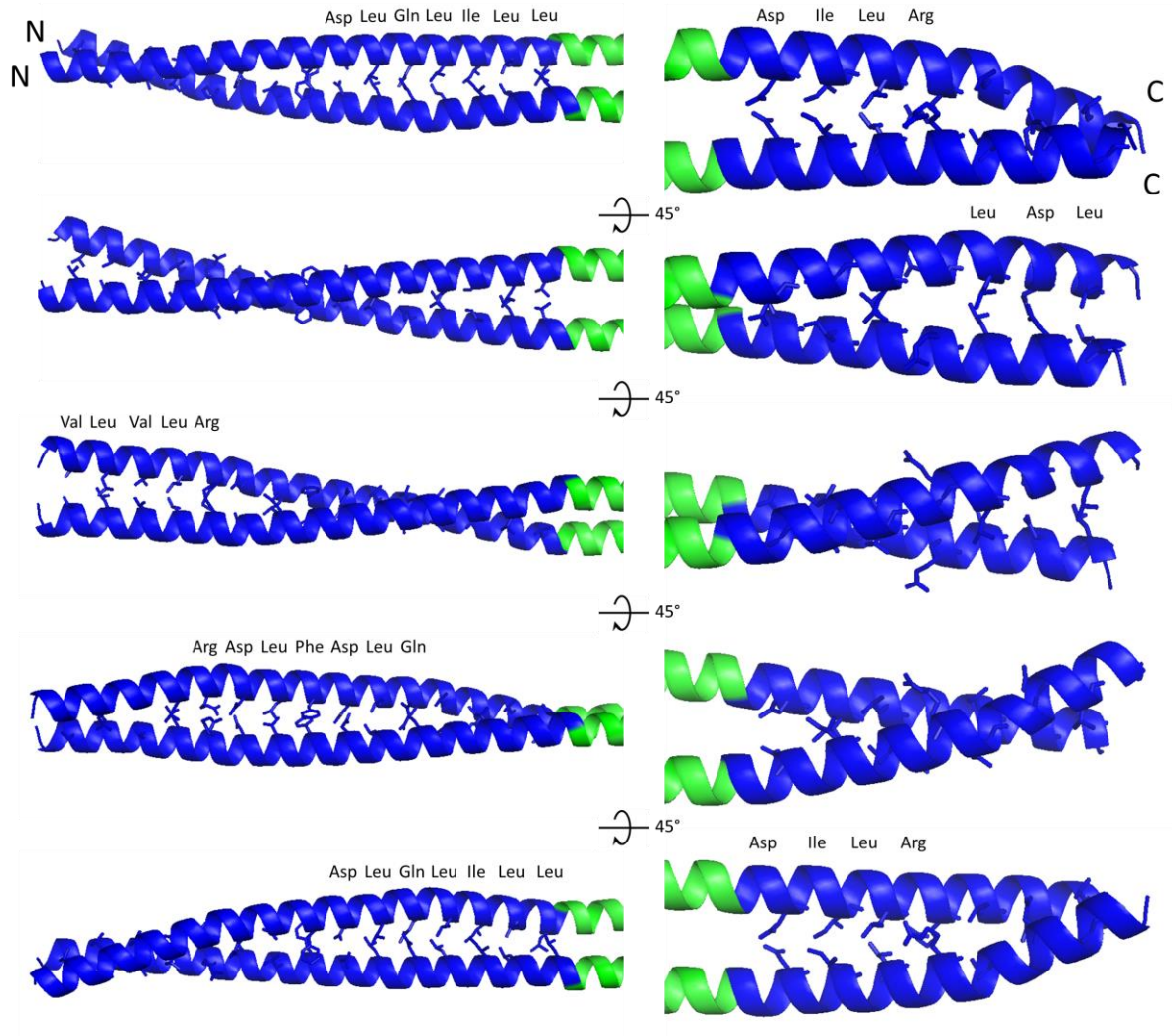
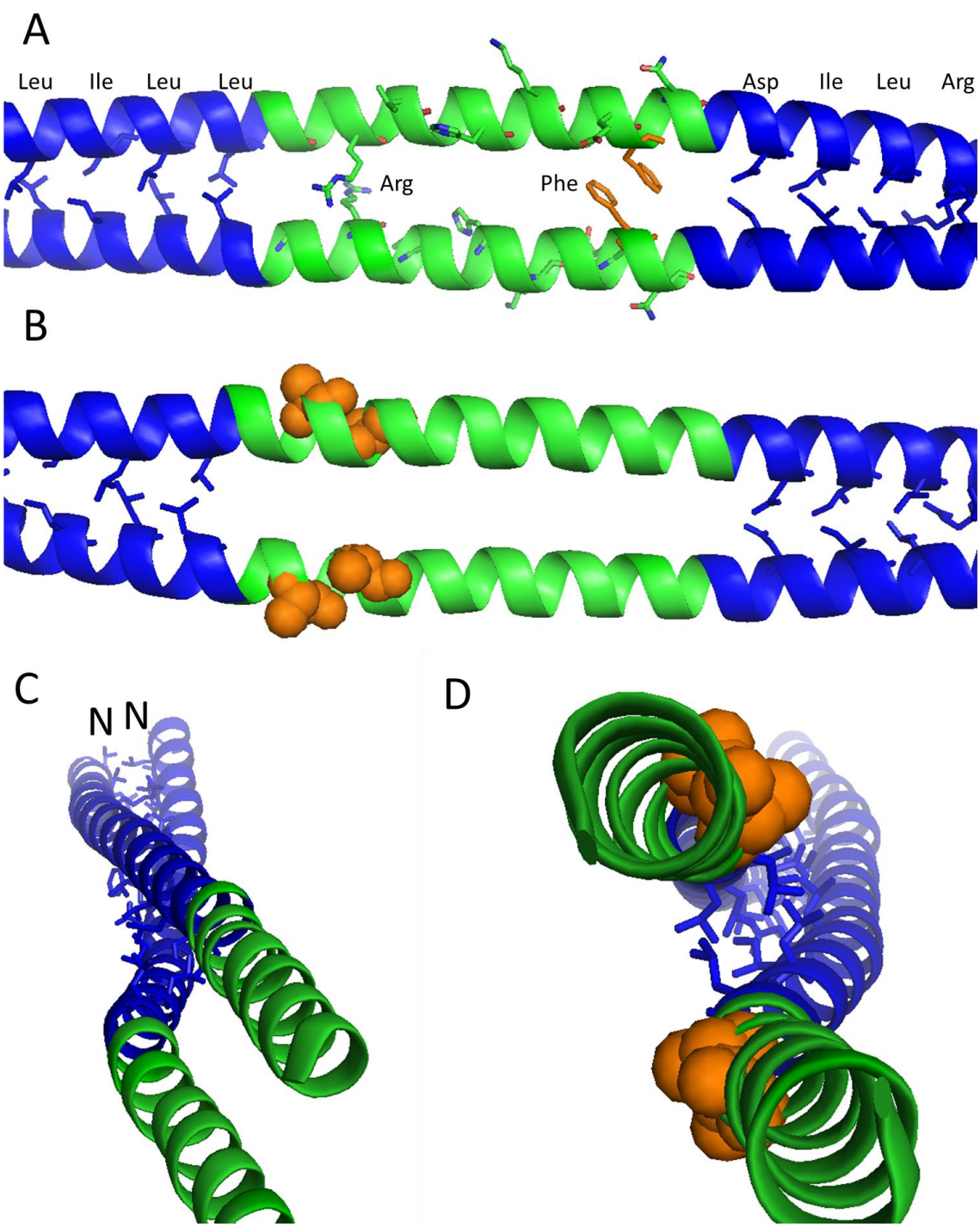


Figure 3.2.7-4: **The structure of NTD$_{31-145}$DM.** The secondary structure is colour coded with blue representing the coiled-coils (amino acids 31-85 and 110-136) separated by a parallel alpha helix in green. Top structure is rotated 90° to obtain the second structure with end on representations.



31HDREVQGLQVKVTKLKQERILDAQRLEEFFTKNQQLREQQKVLHETIKVLEDRLR 85
86 AGLADRAAVTEEHMRKKQQEFENI 109
110 RQQNLKLITELMNERNTLQEENKKLSEQ 137

Figure 3.2.7-5: **Heptad repeat A, D representation of the two coiled coils (31-85 and 110-136) with sequence**. Residues shown as a stick representation in the models are highlighted in cyan in the sequence.



73LHETIKVLEDRLRAGLADRAAVTEEHMRKKQQEFENIRQQNLKLITELMNER 124

Figure 3.2.7-6: **Residues 86-109 form a set of parallel alpha helices with C89A&C92A mutation residues solvent exposed. A)** A, D heptad repeat sequence from the first coiled-coil is continued however residues are no longer in the correct positions for coiled-coil formation shown in C. Arginine 91 is pointing towards the solvent. **B)** Location of the C89A/C92A mutations with alanine residues shown as a space-fill representation. **C)** Visualisation of the change from coiled-coil (blue) to parallel alpha helices (green). **D)** End on view of the C89A/C92A mutations showing their orientation away from the centre of the two parallel helices.

3.2.8 SEC-SAXS of the NTD

SEC-SAXS is a powerful technique that can be used to obtain structural information about ordered and disordered regions of biological samples to a low resolution (Feigin and Svergun, 1989). Here SEC-SAXS is used for determination of the orientation of the alpha helices and gain insights to supplement the crystal structure.

3.2.8.1 Analysis of strep tag addition on protein folding.

Tagged and untagged NTD_{31-145} were analysed by SEC-SAXS to assess the effect of the Strep tag addition on the protein structure. The scattering-based Kratky-plot for the two samples gave similarly shaped curves with both indicating an elongated protein (figure 3.2.8.1A.) Performing a Guinier analysis of the SAXS data showed a slight increase as expected for increasing the length of the protein (figure 3.2.8.1B). Further data analysis using the Pair-distribution function gave D_{MAX}. This value increased by 10 Å, however, the data still gave a P(r) distribution with a steady decay (figure 3.2.8.1C). Collectivly the scattering data indicate the strep-tag did not affect the formation of a long rod shaped protein.

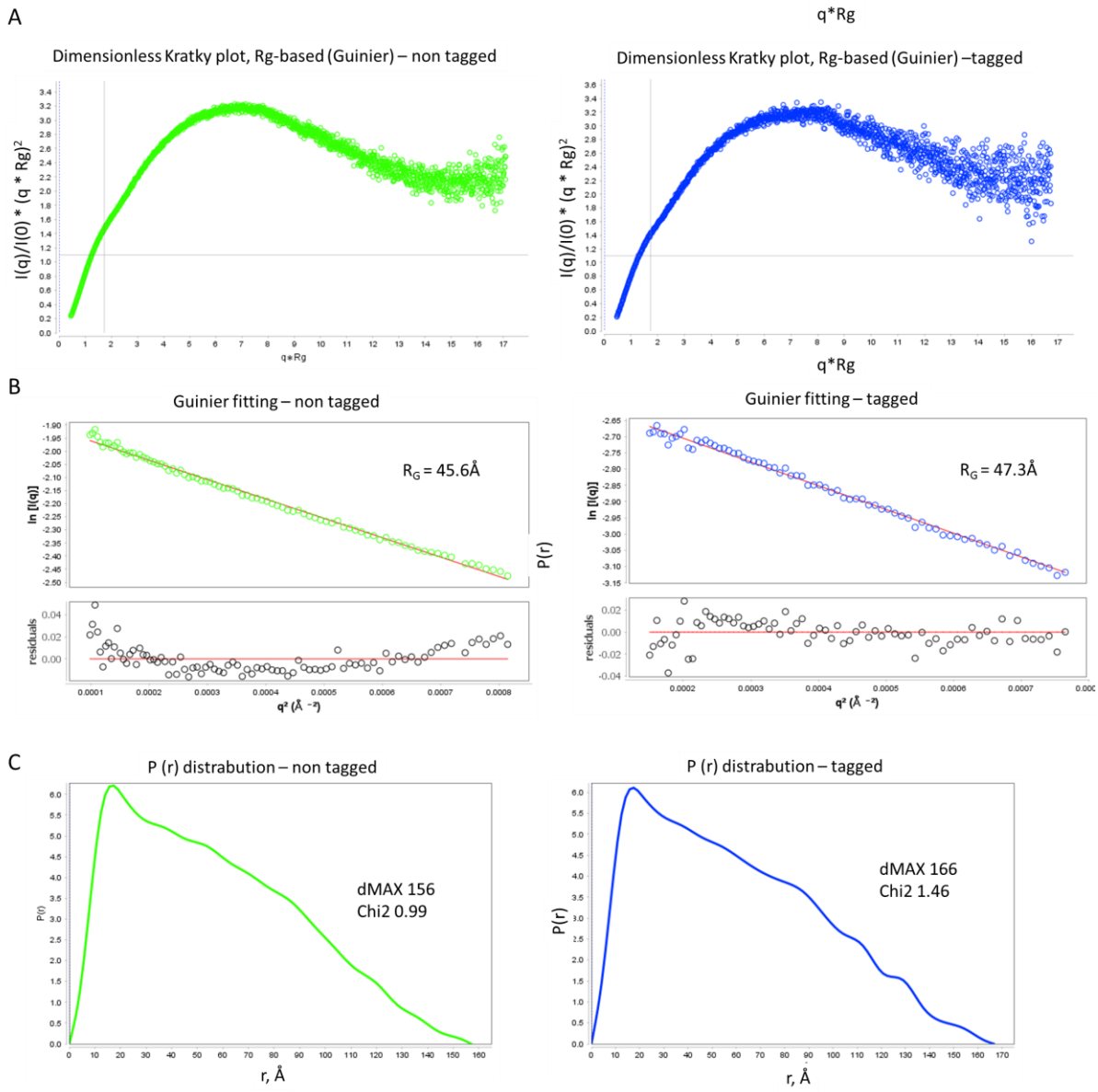


Figure 3.2.8.1: **SEC-SAXS analysis of NTD$_{31\text{-}145}$ vs NTD$_{31\text{-}145}$Strep. A)** Dimensionless Kratky plot for non-tagged NTD$_{31\text{-}145}$, and NTD$_{31\text{-}145}$Strep. **B)** Guinier fitting for non-tagged NTD$_{31\text{-}145}$, and NTD$_{31\text{-}145}$Strep. **C)** P(r) distribution for non-tagged NTD$_{31\text{-}145}$, and NTD$_{31\text{-}145}$Strep.

3.2.8.1.1 The NTD exists as a parallel coiled-coil

Svergun and Koch's article in 2003 depicts a comparison between pair-distribution functions for 5 different shaped particles showing that shape information can easily be obtained from the scattering data (Svergun and Koch, 2003). From the pair distribution function or P(r) distribution a protein envelope can be modelled as this shows all distances between the scattered electrons. Distinctive differences can therefore be seen between rod, globular, and dumbbell shaped proteins from P(r) distribution (figure 3.2.8.1.1) and enable a model to be generated. Addition of bulky protein tags to a single N- or C-terminus of a coiled-coil can determine the orientation of the coil from the change in shape. For a parallel coiled-coil a Y shaped envelope is expected, whereas an anti-parallel coiled coil form a dumbbell shape with a distinctive double peak distribution (Kim et al., 2017). In 2004 four models were proposed for the architecture of the N-terminal dimerisation domain: a straight rod of 17nm, two shorter coils with a flexible region, a 4 helix bundle, and a 'rad50 like' zinc hook (Dubin et al., 2004). To confirm the parallel nature of the domain, in accordance to the crystal structure, SEC-SAXS was performed on an MBP-fused NTD.

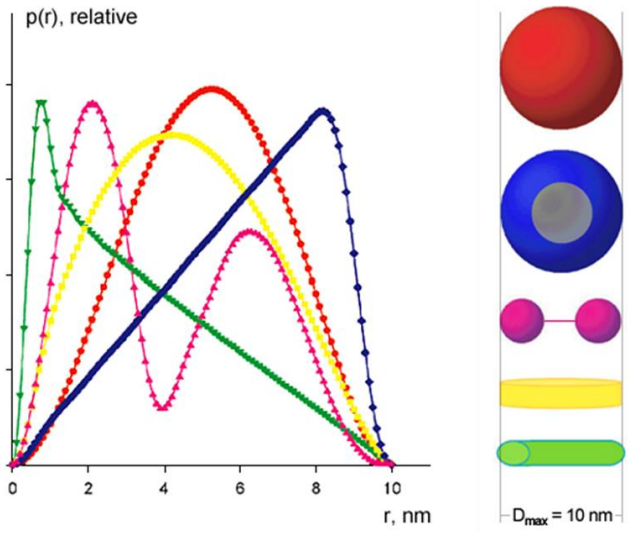


Figure 3.2.8.1.1**: Comparison of P(r) distribution for 5 different shapes.** A sphere in red, doughnut in blue, dumbbell in pink, disk in yellow and a rod in green, all with an identical D_{MAX} of 10 nm are shown. The graph on the LHS shows the P(r) distribution for the 5 shapes. This shows how shape information can be obtained from scattering data, adapted from Svergun and Koch 2003.

3.2.8.1.1.1 Purification of MBP tagged $NTD_{30\text{-}145}$Strep

The MBP tag was already used in purification of NTD constructs, therefore, for an MBP fusion the TEV protease stage was omitted. The soluble protein was purified using sequential Ni-NTA and Strep-Tactin affinity chromatography prior to Superdex 200 16/60 SEC (figure 3.2.8.1.1.1). Fractions containing MBP-$NTD_{30\text{-}145}$Strep were pooled and concentrated for SAXS analysis. MBP was purified as a bi-product from purification of cleaved $NTD_{31\text{-}145}$Strep, for this the MBP was bound in excess to a 2 mL amylose column and purified by a Superdex 75 16/60 SEC column.

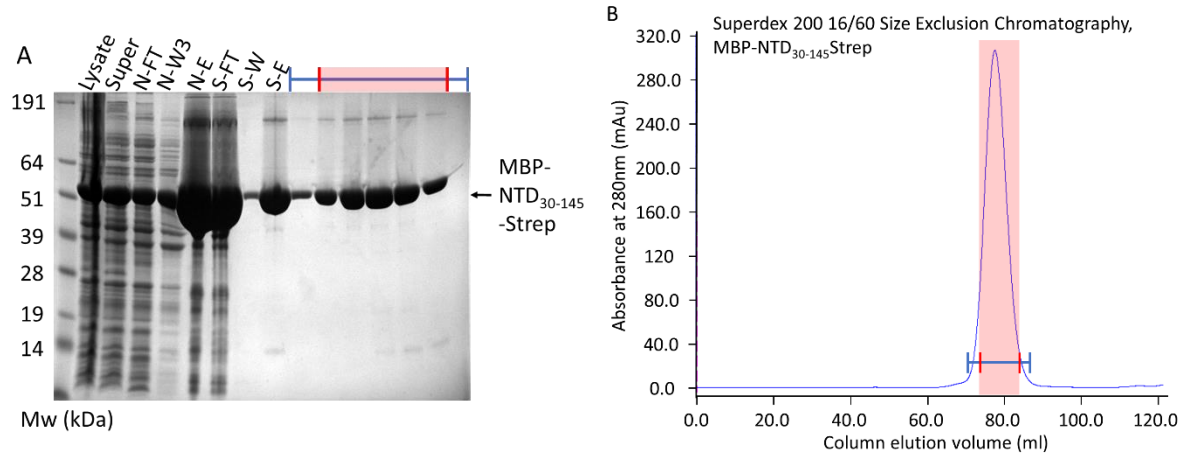


Figure 3.2.8.1.1.1**: Purification of MBP-$NTD_{30\text{-}145}$Strep by Ni-NTA and Strep-Tactin. A)** MBP-$NTD_{30\text{-}145}$Strep was captured by Ni-NTA chromatography, 200 mM imidazole elution fraction (N-E) was bound to a Strep-Tactin column to remove impurities with the Strep elution (S-E) used for gel filtration. Blue lines indicate fractions analysed by SDS-PAGE from gel filtration in B, with the red box indicating pooled fractions for concentration and freezing. **B)** Chromatogram of the Superdex 200 16/60 SEC gel filtration.

3.2.8.1.1.2 SAXS Data analysis using SCÅTTER

SEC-SAXS experiments were performed on B21 at Diamond Light Source individually on $NTD_{30\text{-}145}$Strep, MBP-$NTD_{30\text{-}145}$Strep and MBP to generate small angle scattering data for analysis. Comparisons between the scattering based dimensionless Kratky-plot can give first information about the proteins. MBP has a bell-shaped curve with an inflection at 1.0 which is typical of a globular protein. In contrast $NTD_{30\text{-}145}$Strep has an increasing slope with a shifted and much higher inflection point indicative of a coiled-coil (Putnam et al., 2007) (figure 3.2.8.1.1.2-1). The addition of MBP onto the NTD results in an elongated trace with a similar inflection peak still showing a globular nature to the protein. Guinier fitting was performed to calculate R_G with a larger value for MBP $NTD_{31\text{-}145}$Strep as expected (figure 3.2.8.1.1.2-2). The scattering data were further processed to obtain the P(r) distributions, both MBP and $NTD_{31\text{-}145}$Strep gave P(r) distributions that are indicative of their globular and coiled-coil nature respectively (figure 3.2.8.1.1.2-1B). The single peak for MBP-$NTD_{31\text{-}145}$Strep shows that this does not

form a dumbbell shape but has an appearance between that of the NTD coiled-coil and globular MBP protein.

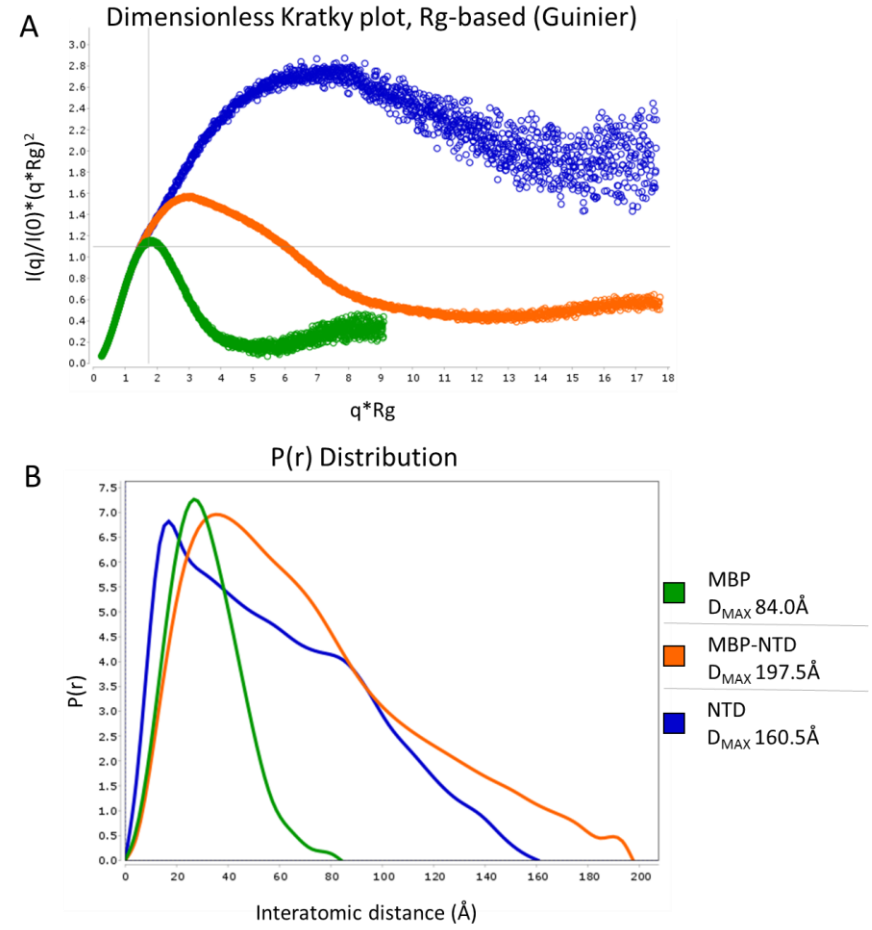


Figure 3.2.8.1.1.2-1: **SEC-SAXS analysis of NTD$_{30-145}$Strep fused to MBP. A)** The Kratky plot, **B)** the P(r) distribution. Both indicate that the NTD is a long molecule and show the globular nature of MBP. From the traces we can see that the combined MBP-NTD has an elongated globular appearance indicating a parallel coiled-coil.

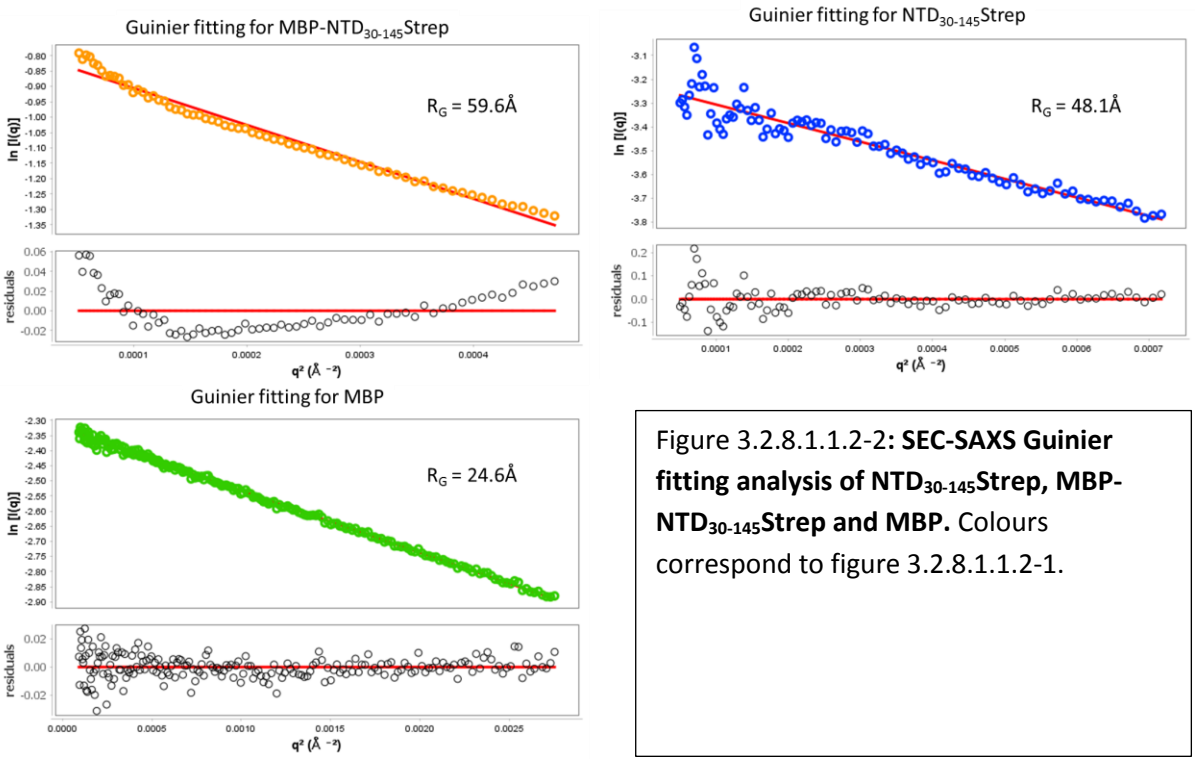


Figure 3.2.8.1.1.2-2: **SEC-SAXS Guinier fitting analysis of NTD$_{30-145}$Strep, MBP-NTD$_{30-145}$Strep and MBP.** Colours correspond to figure 3.2.8.1.1.2-1.

3.2.8.1.1.3 SAXS model comparisons

Models to fit the SAXS data were generated through SCÅTTER using the ATSAS package program DAMMIF and DAMAVA. The resultant DAMFIT files from averaging were visualised using PyMol and aligned manually. To assess the quality of the data the model to fit the SAXS data for MBP were aligned with its known structure (3hpi) with no obvious discrepancies (figure 3.2.8.1.1.3A).

The MBP tag on the N-terminus of the $NTD_{30\text{-}145}$Strep construct will allow the location of this end to be determined. Alignment of the two models for $NTD_{30\text{-}145}$ along with MBP-$NTD_{30\text{-}145}$ show that the MBP-$NTD_{30\text{-}145}$Strep has a much larger envelope and a non-symmetrical appearance (figure 3.2.8.1.1.3B). The protruding section within this model is sufficient to align the $NTD_{30\text{-}145}$Strep model. Further alignment of the model to fit the MBP SAXS data was only possible within one end of the MBP-$NTD_{30\text{-}145}$ SAXS envelope. However, the large space within one side of the model was sufficient for two MBP envelopes to be placed (figure 3.2.8.1.1.3C). From this it can be concluded that the CtIP NTD forms a parallel dimer in solution. As DAMAVER is known to average out flexibility, individual models were compared with no drastic differences to the averaged model.

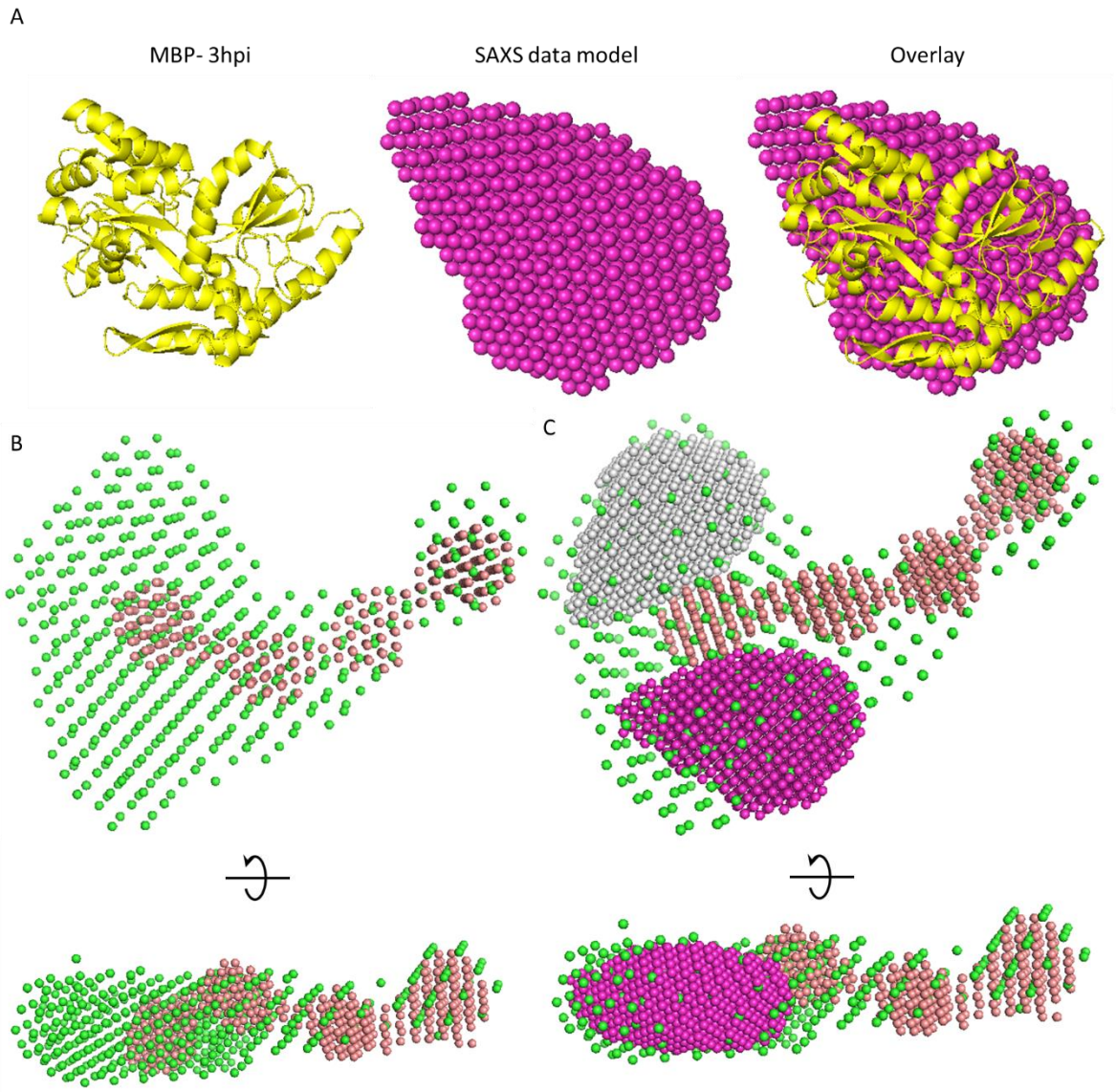


Figure 3.2.8.1.1.3: **DAMFIT models to fit the SAXS data of MBP, NTD_{30-145}Strep and MBP-NTD_{30-145}Strep. A)** Left to right, PDB model 3hpi of MBP, model to fit SAXS data, overlay of the two models. **B)** Overlay of NTD_{30-145}Strep with MBP-NTD_{30-145}Strep, adequate room for the coil-coil to fit within the protrusion of the fusion protein is observed. **C)** Docking of two models to fit the MBP SAXS data within one side only of the MBP-NTD_{30-145}Strep SAXS data model. From this it can be concluded that the NTD forms a parallel coiled coil.

3.2.8.1.2 Assessing the architecture of the coiled coil

As shown by secondary structure prediction (Dubin et al., 2004) and the determined crystal structure, the CtIP coiled-coil domain can be sub divided into two regions separated at the zinc binding motif. Performing a Guinier analysis of initial slope of SEC-SAXS data allows calculation of R_G. Reducing the length (L) of a straight coiled-coil by half should reduce R_G by half. This assumes the cross-sectional radius of gyration (R_C) remains constant determined by equation 3.2.8.1.2.

Equation 3.2.8.1.2

$$R_g{}^2 - R_c{}^2 = \frac{L^2}{12}$$

To confirm the elongated coiled coil is the correct model the NTD was separated at the zinc binding site and two individual sub domains were cloned, expressed and purified for scattering analysis. Calculating R_g for the two sub domains verses $NTD_{31\text{-}145}$Strep will help elucidate the quaternary structure of the coiled-coils.

3.2.8.1.2.1 Construct design and protein purification

Constructs spanning from CtIP residues 31-88 ($NTD_{31\text{-}88}$Strep) and 93-145 ($NTD_{93\text{-}145}$Strep) were cloned into pMAT11 with an additional C-terminal Strep-tag. Positive clones were confirmed by DNA sequencing by the Department of Biochemistry's DNA Sequencing Facility prior to protein expression using the Rosetta™ 2 (DE3) cell line.

Two litres of cells expressing $NTD_{93\text{-}145}$Strep and four litres of cells expressing $NTD_{31\text{-}88}$Strep were used for each purification using the standard C-terminal strep tagged purification strategy. For SDS-PAGE analysis 4-12% polyacrylamide gels were used to enable visualisation. Fractions containing NTD protein were pooled and concentrated to 3.69 mg/mL for $NTD_{31\text{-}88}$Strep and 5.86 mg/mL for $NTD_{93\text{-}145}$Strep (figure 3.2.8.1.2.1).

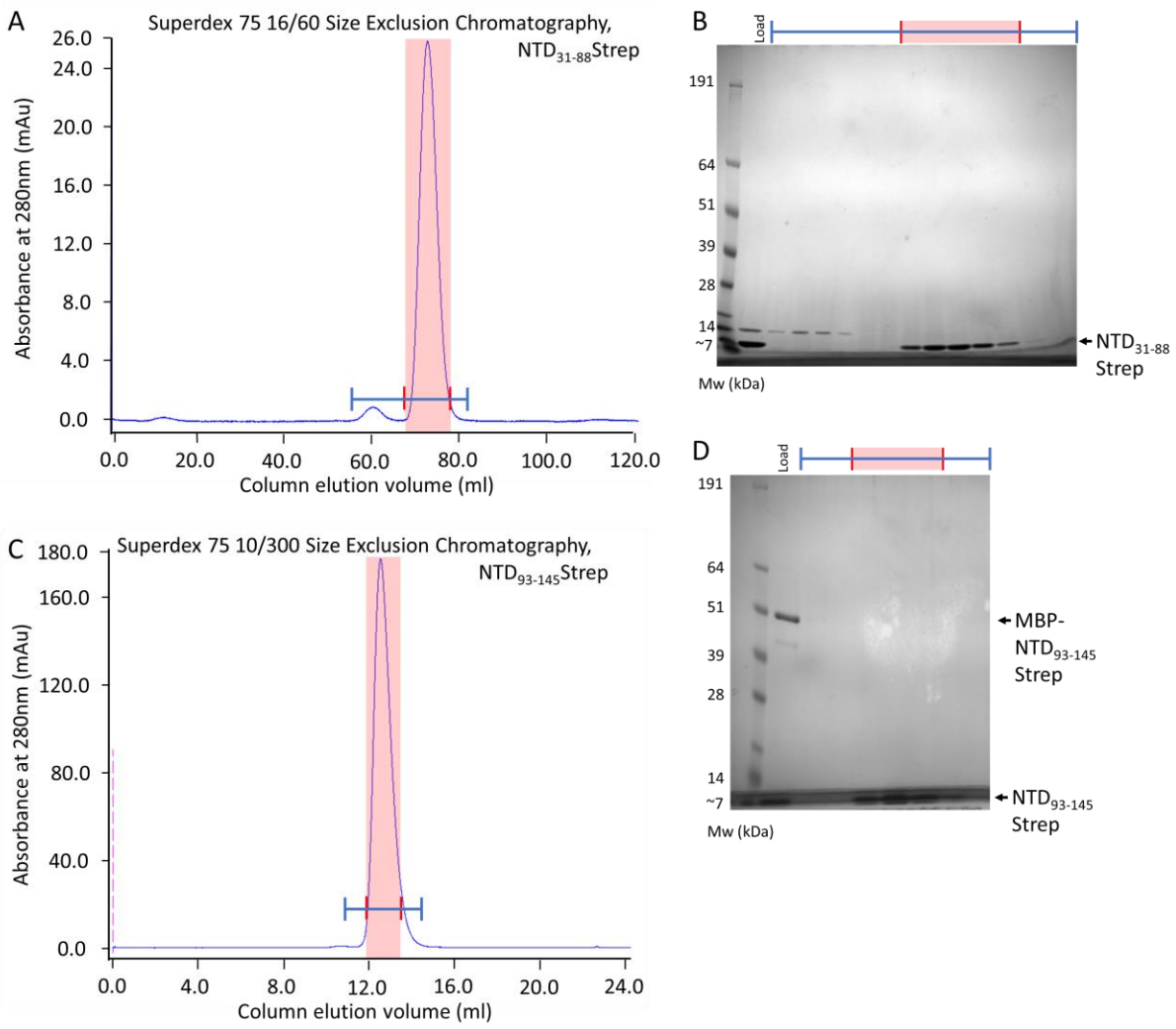


Figure 3.2.8.1.2.1: **Final stage purification of NTD_{31-88}Strep and NTD_{93-145}Strep.** Proteins were purified using the standard N-&C-terminal tagged purification procedure. **A)** SEC chromatogram for NTD_{31-88}Strep, fractions were analysed by SDS-PAGE in B. **B)** Protein applied to the SEC column (load) precedes elution fractions. First peak contains a protein of 14 kDa with the main peak containing NTD_{31-88}Strep. Fractions in red were pooled for concentration and experimental analysis. **C)** SEC chromatogram for NTD_{93-145}Strep analysed by SDS-PAGE in D. **D)** Protein applied to the SEC column (load) precedes elution fractions. Fractions in red were pooled for concentration and experimental analysis.

3.2.8.1.2.2 SAXS data analysis using SCÅTTER

SEC-SAXS experiments were performed on B21 at Diamond Light Source individually on NTD_{31-88}Strep, NTD_{93-145}Strep, NTD_{31-145}DM-Strep and NTD_{31-145}Strep to generate small angle scattering data for analysis. Comparisons between the scatter-bases Kratky-plots can give a first indication of any change in length and the coiled nature of the sub domains proteins. Both sub domains maintained an inflection point greater than 1 indicative of coiled-coil protein. The sub domains however do have a lower inflection point than NTD_{31-145} indicating a shorter coiled-coil protein (figure 3.2.8.1.2.2-1). Comparison between the wild type and double mutant proteins show minimal differences with overlaying traces.

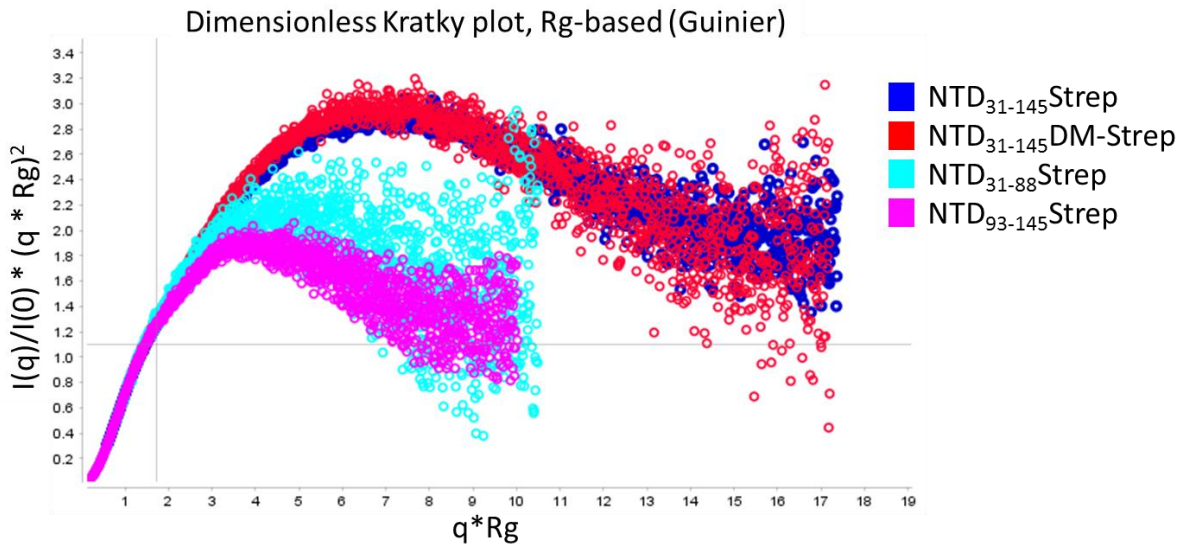


Figure 3.2.8.1.2.2-1**: Assessing the architecture of the NTD SEC-SAXS Dimensionless Kratky plot.** The graph shows the Kratky plot for the two halves of the NTD, NTD_{31-88}Strep and NTD_{93-145}Strep verses NTD_{31-145}Strep with and without the zinc-binding site (NTD_{31-145}DM-Strep). The reduction in peak height compared to NTD_{31-145}Strep for the two shorter constructs indicates a reduced length of these proteins. The difference between the DM and wild type protein is minimal, however the higher peak for NTD_{31-145}DM-Strep indicates a less compact protein.

Guinier fitting of the scattering data to obtain R_G and further P(r) distribution and D_{MAX} values for the four proteins was performed. With equation 3.2.8.1.2 comparisons between these values can provide an insight into the domain architecture. The NTD_{31-88} (R_G 30.0) and NTD_{93-145} (R_G 28.7) R_G values are not the same as the NTD_{31-145} (R_G 48.5 or 48.1 for the DM) (figure 3.2.8.1.2.2-2A/B). This provides further evidence that the domain forms an elongated coiled-coil, as reduction in the length of the protein has reduced R_G.

Comparison between these data and those presented in figure 3.2.8.1.2.2-2C show that there is variability within the R_G and D_{MAX} between the two independent SEC-SAXS data collections. The same column type was used for both experiments, with a broadening of the SEC peak noted for the second data collection.

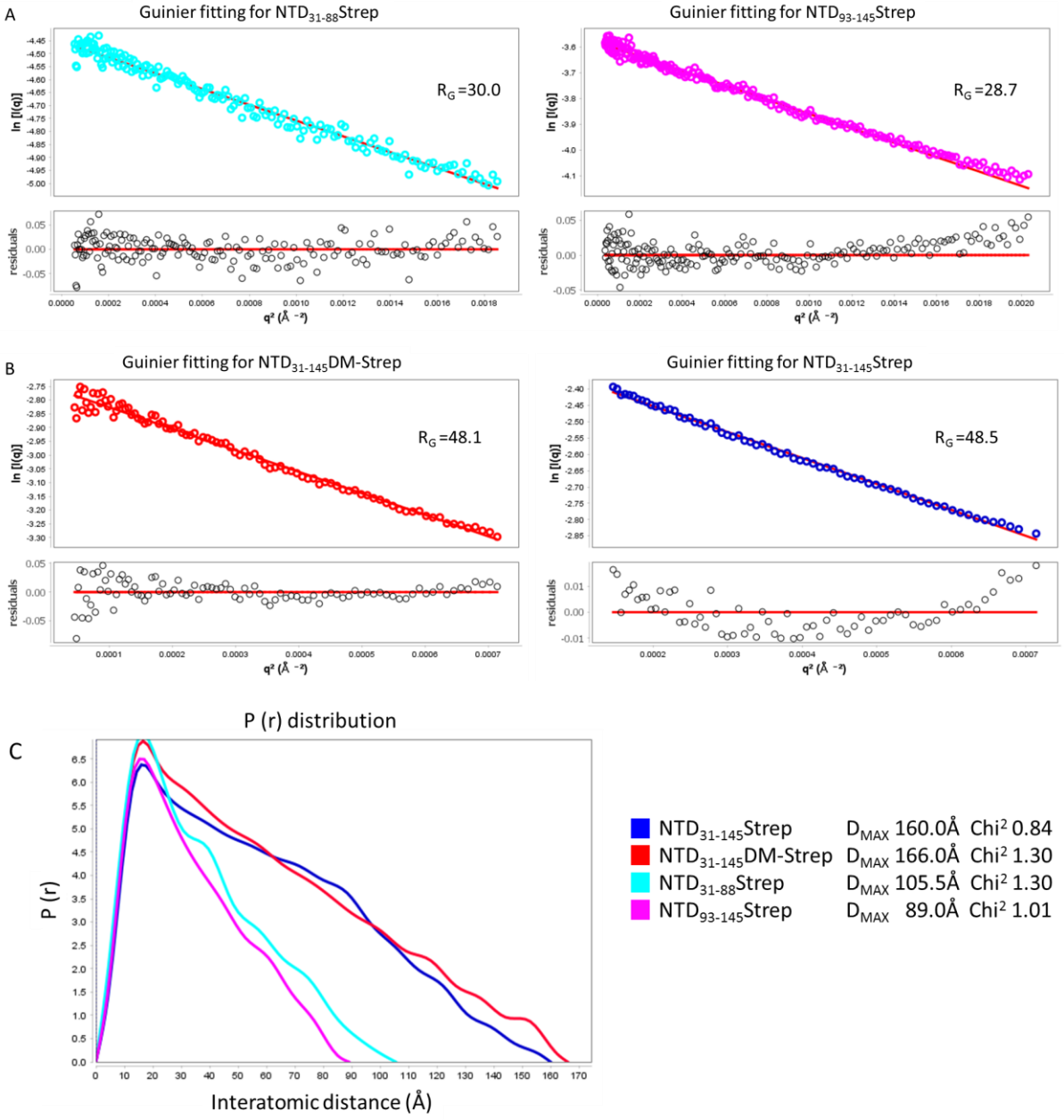


Figure 3.2.8.1.2.2-2: **SEC-SAXS analysis of NTD_{31-88}Strep, NTD_{93-145}Strep, NTD_{31-145}Strep and NTD_{31-145}DM-Strep, Guinier fitting and P(r) distributions. A)** Guinier fitting for NTD_{31-88}Strep (LHS) and NTD_{93-145}Strep (RHS) with R_G displayed in Å. **B)** Guinier fitting for NTD_{31-145}DM-Strep (LHS) and NTD_{31-145}Strep (RHS) with R_G displayed in Å. **C)** P(r) distribution with D_{MAX} showed. Colours correspond in figure 3.2.8.1.2.2-1.

The P(r) distributions show that the proteins decay towards 0 which is typical for rod structures. However, the bumpiness of the decay highlights the data collection or processing is not perfect. Bumpiness within the P(r) destitution is also apparent in comparison between with and without the Strep tag indicating this may influence the distribution. Nevertheless, model generation for NTD_{31-145}DM-Strep shows that the crystal structure fits into the model to fit the SEC-SAXS data with extra space within the envelope for the Strep tag at the C-termini of the protein (figure 3.2.8.1.2.2-3). Determination of D_{MAX} and R_G from the NTD_{31-145} crystal structure using SCÅTTER gives a value of 164 Å and 45 Å respectively which are similar to the values obtained by SAXS.

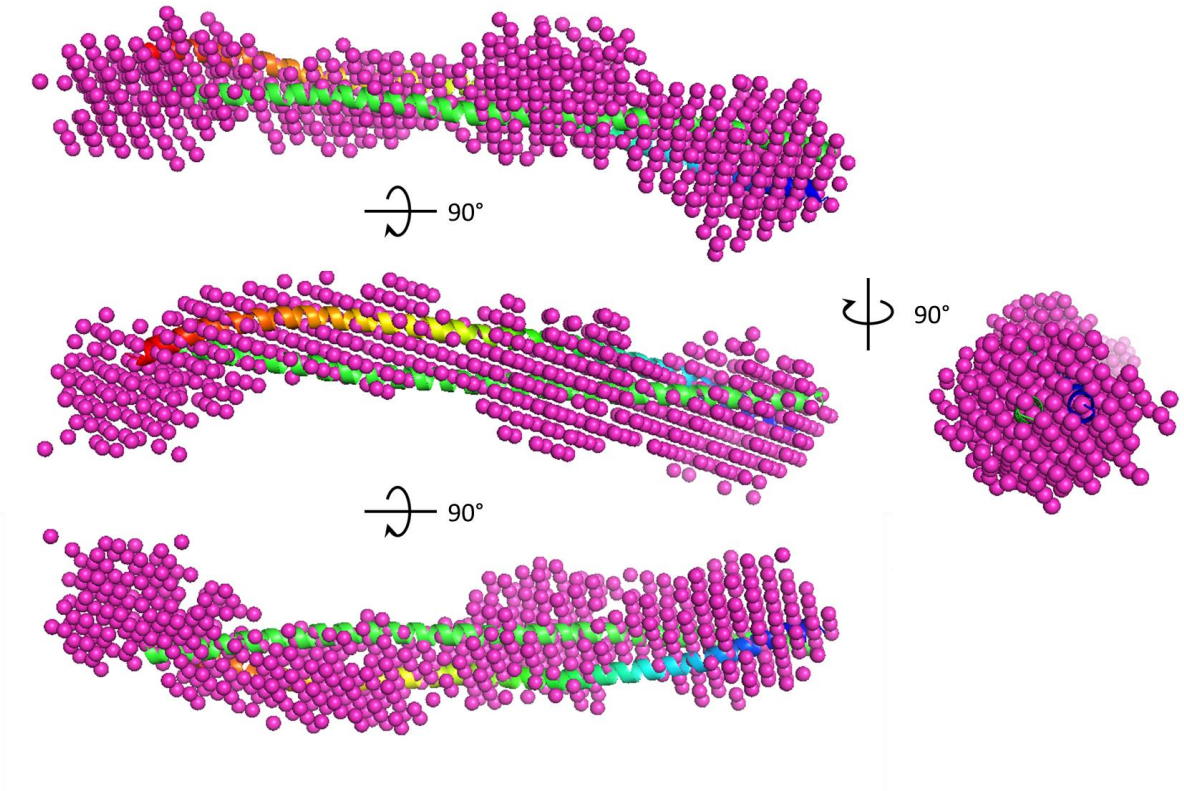


Figure 3.2.8.1.2.2-3: **Overlay of NTD_{31-145}DM-Strep crystal structure with the model to fit the SEC-SAXS data.** Figure was made manually in PyMol with crystal structure fitting within the SAXS envelope. Crystal structure is coloured from Blue to red from N- to C-terminus.

3.2.9 Structural analysis of the tetrameric NTD.

The NTD contains multiple structural elements of known functional importance. However, the global architecture of the tetrameric domain is unknown. Efforts to crystallise CtIP residues 18-145 were broadly unsuccessful with no tractable conditions found. Alternative structural characterisation methods were therefore sought.

3.2.9.1 Cryo EM analysis of NTD_{18-145}

Cryo-EM avoids the need for protein crystals but is typically restricted to large protein domains and complexes (>100 kDa). The relatively small size of the domain impedes effective analysis by this technique. Therefore, an MBP fusion of CtIP NTD was expressed giving a tetrameric protein with 4 large globular MBP molecules to aid particle picking (figure 3.2.9.1). As residues 1-18 are predicted flexible (Sun, 2013) to increase the likelihood of selecting tetrameric particles the construct spanned from residues 18-145 (MBP-NTD_{18-145}) at a molecular weight of 242 kDa.

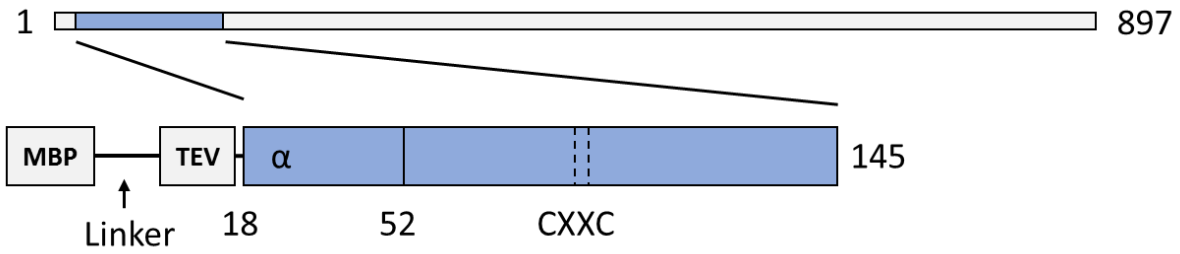


Figure 3.2.9.1: **Schematic drawing of NTD_{18-145}.** NTD_{18-145} is indicated in blue with the TEV cleavable MBP tag. The tetramerisation domain (α) alongside the boundaries for its structure (18-52, PDB 4D2H) and the zinc binding domain (CXXC) are also shown.

3.2.9.1.1 Recombinant protein expression and purification, MBP-NTD_{18-145}

The MBP-NTD_{18-145} construct was expressed in two litres of Rosetta™ 2 (DE3) bacteria. The soluble protein was purified using sequential affinity chromatography: Ni-NTA resin, amylose resin and a Hi-Trap Ni-NTA resin column, to give a concentrated fraction of fusion protein. Final polishing was performed using a 16/60 Superdex 200 SEC column (figure 3.2.9.1.1). Fractions were analysed by SDS-PAGE with protein concentration of each fraction determined by absorbance at 280 nm. The peak fraction (figure 3.2.9.1.1) from the Superdex 200 16/60 SEC contained the most protein at 3.1 mg/mL and was used for preparation of Cryo-EM grids.

All grid manipulations were performed by Dr Neil Rzechorzek using grids from Batch No.190698, QuantiFoil R 1.2/1.3 Cu300 and 3 µL of fresh protein per grid.

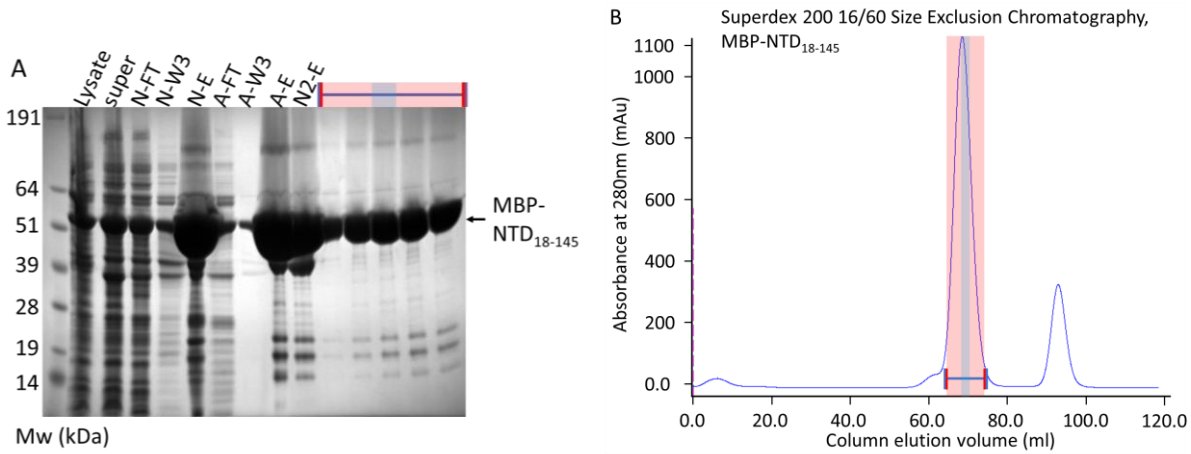


Figure 3.2.9.1.1**: Purification of MBP-NTD$_{18\text{-}145}$. A)** The soluble protein was applied to a Ni-NTA resin, amylose resin and a His-Trap Ni-NTA column without TEV cleavage. The elution from the His-Trap Ni-NTA column was polished by Superdex 200 16/60 SEC **B)** Chromatogram showing SEC of MBP-NTD$_{18\text{-}145}$: blue lines indicate fractions analysed by SDS-PAGE in A, with red fractions pooled for concentration and freezing. The fraction highlighted in blue was used for Cryo-EM.

3.2.9.1.2 Cryo-EM data collection

Grids containing MBP-NTD$_{18\text{-}145}$ were analysed using the Talos Arctica™ microscope controlled by Dr Neil Rzechorzek at the Department of Biochemistry Cryo-EM facility. An ice gradient was present across each grid analysed, however ice thickness limited square selection (figure 3.2.9.1.2-1). Squares were screened across each grid, however in many cases cracked ice inhibited image collection.

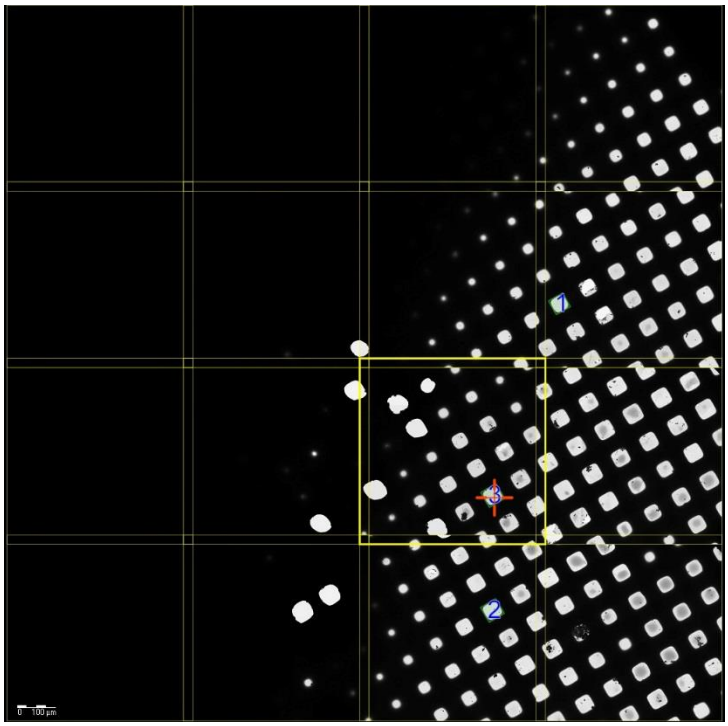


Figure 3.2.9.1.2-1: **MBP-NTD$_{18\text{-}145}$ Atlas image of grid indicating 3 squares imaged.**

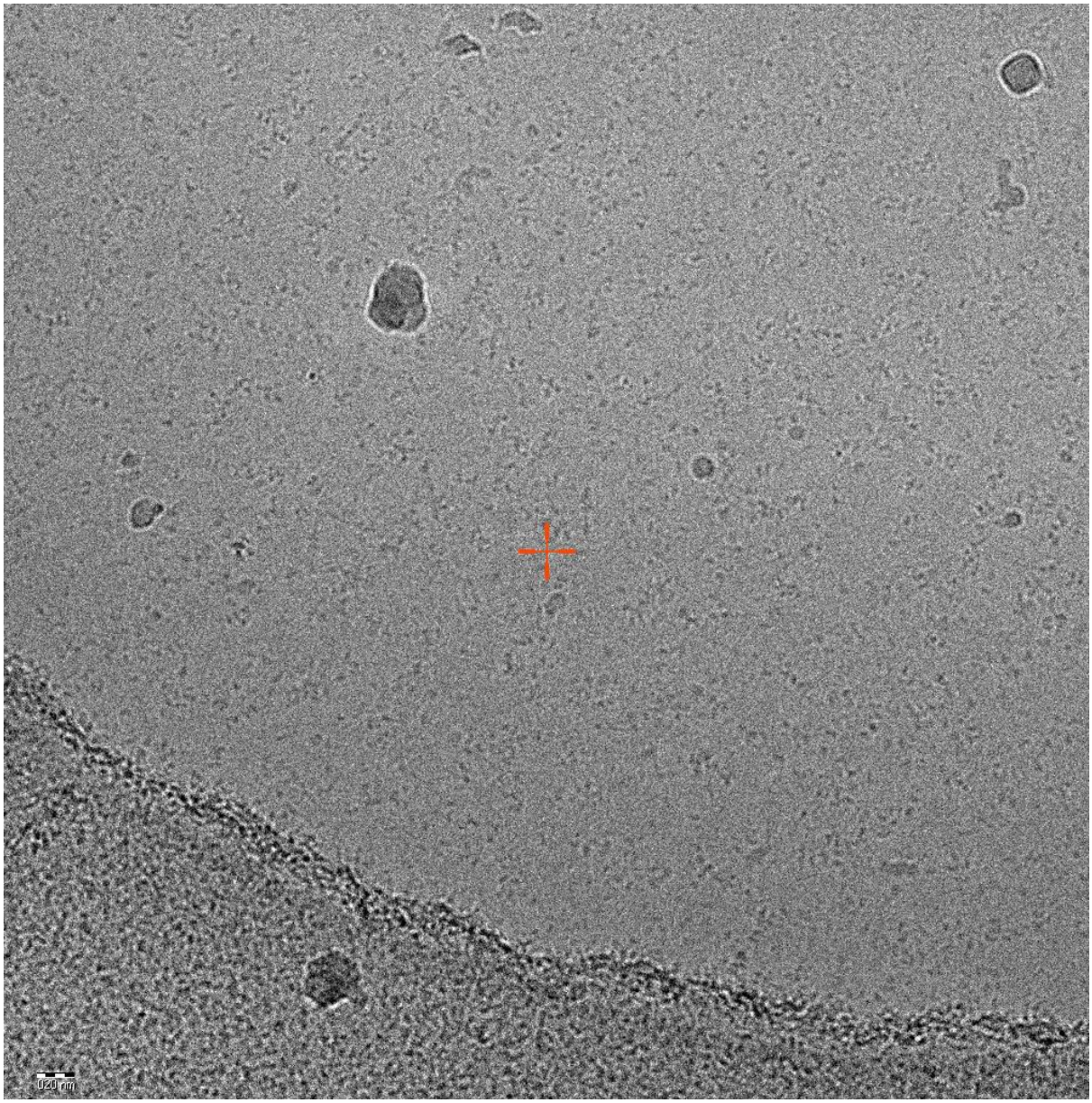


Figure 3.2.9.1.2-2**: MBP-NTD$_{18-145}$, Representative Talos Arctica™ microscope image.** Many black particles can be seen across the grid, but a large proportion of protein is aggregated at the carbon edge. Particles are insufficiently clustered to determine tetramerisation with optimisation required.

MBP was assumed to be the small black spots on the image (figure 3.2.9.1.2-2). Over the 4 grids imaged the 12 holes that were sampled gave very similar results: ice thickness made it difficult to select intact squares, with very few in close proximity. The high particle density and the flexibility of the linker region between the NTD and MBP meant that tetrameric particles could not be picked. Construct, protein concentration, grid preparation and data collection optimisation will be necessary to allow particle picking leading to a future structure determination.

3.2.9.2 Small angle x-ray scattering of NTD_{18-145}

Cryo-EM analysis of the NTD has yet to be successful. However, with crystal structures for CtIP residues 18-52 (4D2H) and 31-136 now available, alongside SAXS analysis confirming the parallel nature, a combined model of CtIP residues 18-136 can be generated. To assess the accuracy of such a model SEC-SAXS was performed on the cleaved NTD_{18-145} construct.

3.2.9.2.1 Protein purification of NTD_{18-145}

Initial purification of NTD_{18-145} was through the standard procedure for non-Strep tagged protein with sequential Ni-NTA capture, TEV cleavage and amylose re-capture of MBP. To improve the NTD_{18-145} homogeneity the intrinsic affinity to the Ni-NTA resin was exploited to give a high purity sample for Superdex 200 16/60 SEC (figure 3.2.9.2.1). Fractions by SDS-PAGE that contained NTD_{18-145} were concentrated to 4.42 mg/mL for SAXS analysis (figure 3.2.9.2.1).

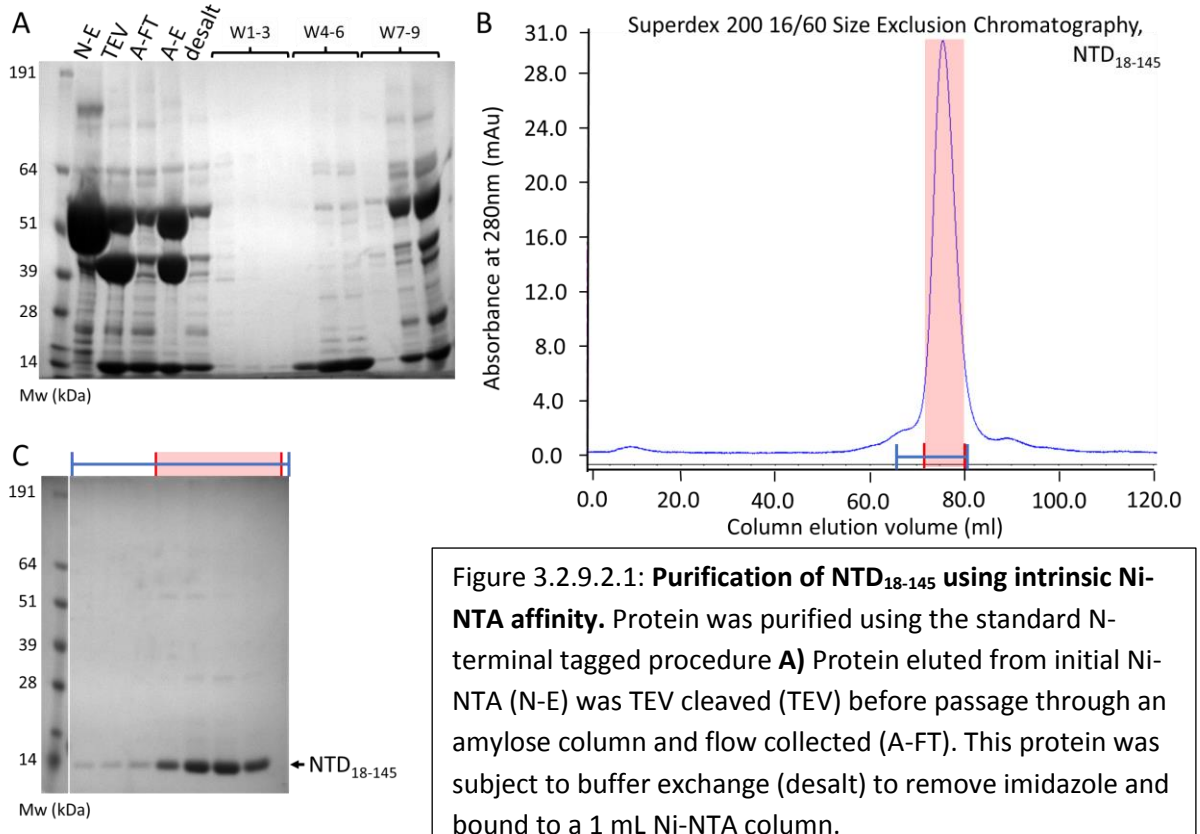


Figure 3.2.9.2.1: **Purification of NTD_{18-145} using intrinsic Ni-NTA affinity.** Protein was purified using the standard N-terminal tagged procedure **A)** Protein eluted from initial Ni-NTA (N-E) was TEV cleaved (TEV) before passage through an amylose column and flow collected (A-FT). This protein was subject to buffer exchange (desalt) to remove imidazole and bound to a 1 mL Ni-NTA column.

A cont.) This column was washed with standard buffer (W1-3) before 20 mM imidazole with standard buffer wash (W4-6) and a final 200 mM imidazole with standard buffer wash (W7-9). W4-6 was pooled for SEC in B. **B)** Superdex 200 16/60 SEC with fractions analysed by SDS-PAGE in C. **C)** analysis of SEC fractions with NTD_{18-145} indicated. Fractions in red were pooled and concentrated for analysis.

3.2.9.2.2 Model generation of CtIP residues 18-145

Analysis of the SEC-SAXS data for NTD_{18-145} gave a value for R_G almost double that of NTD_{31-145} as expected indicating an elongated structure was retained (figure 3.2.9.2.2-1). The data were further processed to give a large D_{MAX} value further supporting an elongated structure. The data however has imperfections shown in the P(r) distribution from a linear decay for a coiled-coil. Generation of an *ab initio* model using DAMMIF therefore gave some discrepancies from a straight coiled-coil with a better sample at higher concentration required (figure 3.2.9.2.2-1).

Combination of the two structures allows visualisation of the 18-136 CtIP structure from the overlapping 16 amino acids between the two structures (figure 3.2.9.2.2-2). The coils from NTD_{31-145} extend into the tetramerisation domain; this further supports the accuracy of the NTD_{31-145}DM-Strep crystal structure as elongation of these coils past residues 31 would form the tetramerisation domain as expected (figure 3.2.9.2.2-2). Generation of SAXS data from the combined structures using SCÅTTER gives a D_{MAX} value of 341 Å, which is within 10% of the SAXS data value (figure 3.2.9.2.2-1), and an R_G value of 83.0 Å, which is within 2% of the SAXS value (figure 3.2.9.2.2-1), indicating the model and SAXS initial processing is likely to be correct.

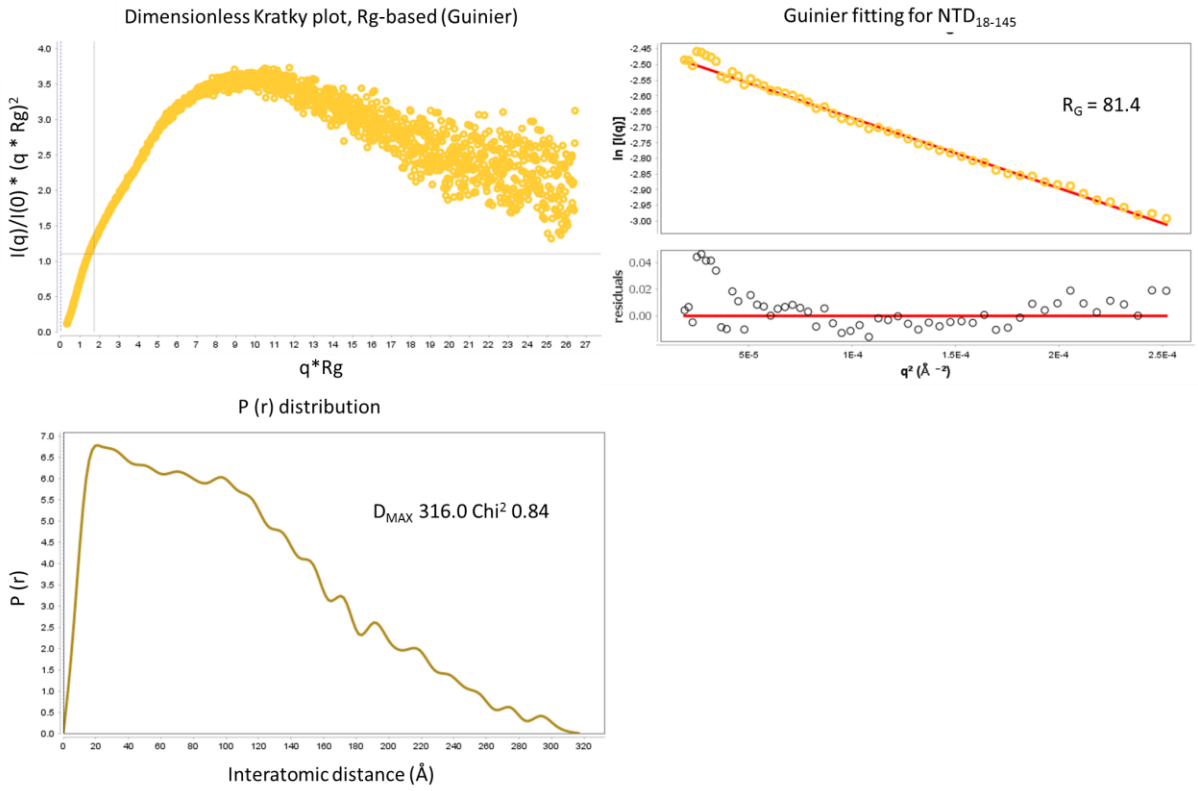


Figure 3.2.9.2.2-1: **SEC-SAXS analysis of NTD_{18-145}**. Going clockwise, first panel is the dimensionless Kratky plot followed by Guinier fitting with R_G displayed. The final panel is the P(r) distribution with the D_{MAX} value displayed.

Despite the imperfections in the P(r) distribution, a model was obtained to fit the SEC-SAXS data for NTD_{18-145}. The combined crystal structure fits within this envelope with both displaying an elongated structure (figure 3.2.9.2.2-2).

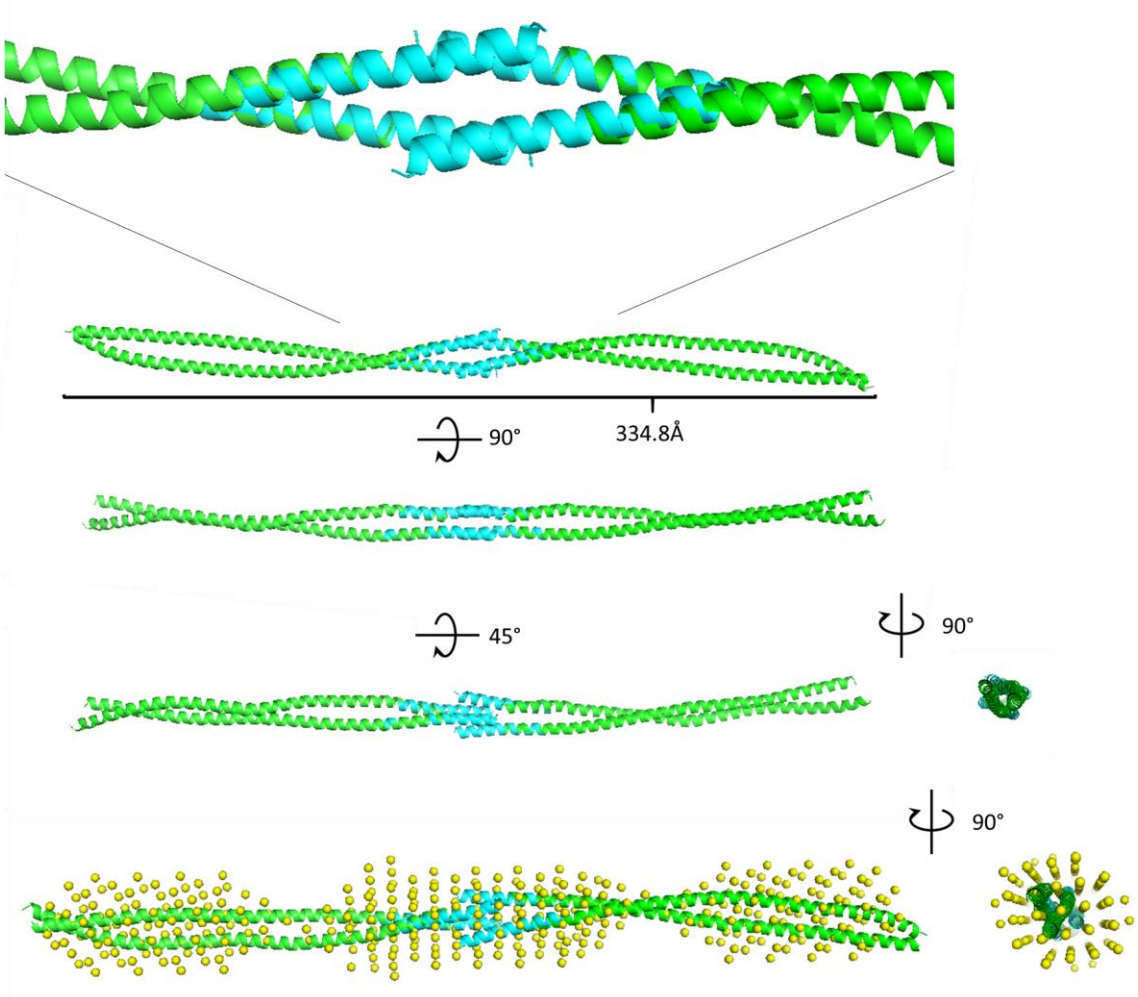


Figure 3.2.9.2.2-2: **Alignment of the CtIP 18-52 (4D2H) and CtIP 31-145 crystal structures overlaid with SAXS envelope for CtIP NTD_{18-145}.** Top structure shows a zoomed image of the alignment between the 16 overlapping residues between the two structures made in PyMol. Next three images are rotations of the aligned CtIP crystal structures with side end view shown. Last panel shows overlaid model to fit NTD_{18-145} SAXS data with the CtIP 18-145 crystal structures with side end view made in PyMol.

3.2.10 Conclusions and future work.

The collective structural studies of the N-terminal region of human CtIP have confirmed the formation of a well-structured dimer which aligns to the previous tetrameric structure (PDB 4D2H). X-ray crystallography and SEC-SAXS have been used to investigate this structure with converging results showing the formation of a parallel dimeric coiled-coil adjacent to the tetramerisation motif.

Formation of crystals for the N-terminal region was optimised through construct design, changes to the purification procedure and crystallisation buffer conditions. A magnitude of different techniques including, seeding, additive screens, changes in temperature, and optimisation of crystallisation conditions improved the resolution from 36 Å to 8.0 Å with many changes in crystal form. This resolution however was insufficient for structure determination by molecular replacement. Expression as SeMet-labelled protein failed to produce protein crystals. Further techniques including Surface Entropy Minimisation (SEM) gave salt crystals which limited further gains in diffraction resolution.

Through a long and challenging crystallographic study the structure of the predicted coiled-coil domain spanning CtIP residues 31-136 with two cysteine-to-alanine mutations has been determined. This shows extension of the alpha helices from the tetramerisation domain into a coiled-coil. The heptad repeat for the coiled-coil formation discontinues at the zinc binding site with formation of parallel alpha helices. The only contact between these two alpha helices is that of two phenylalanine residues with their side chains forming a π-stack of which likely mediates the re-start of the coiled-coil within this structure. Analysis of the alanine mutations (C89A/C92A) show the R' groups of the amino acids point towards the solvent rather than centre of the helix pair. The alanine mutations therefore likely encourage the alpha helical formation with a local change in conformation. However, SEC-SAXS shows minimal differences between the wild type and mutant protein with a model to fit the SEC-SAXS data enclosing the crystal structure. This indicates any change in structure to be minimal around this region. A single point missense mutation in CtIP NTD of arginine at position 100 to tryptophan has been shown to be associated with primordial dwarfism (Shaheen et al., 2014). Visualisation of this amino acid in the structure shows that it is solvent exposed and not involved in formation of the coiled-coil. This mutation therefore is less likely to disrupt the structure and is more likely involved with a protein-protein interaction.

The complexity of the Morpheus screen and its 3-dimensional nature makes it difficult to pinpoint buffer components that encourage crystallisation. From the use of matrix micro seeding a vast increase in crystal numbers was seen, indicating that formation of nuclei is a large issue for the protein. So far wild type protein has yet to crystallise in the Morpheus screen. Matrix micro-seeding of NTD_{31-}

$_{145}$DM-Strep micro-crystals into $NTD_{31\text{-}145}$Strep protein crystallisation trials have also yet to be successful.

SEC-SAXS enabled further biophysical characterisation of the NTD with an MBP fusion confirming the region forms a parallel coiled-coil in solution. The architecture of this domain was further explored through SEC-SAXS analysis of the two sub regions (CtIP residues 31-88 and 93-145) that are separated by the zinc-binding domain. These two studies provided evidence towards a long coiled-coil of residues 31-136 shown by crystallography and ruled out the formation of a Rad50 zinc hook structure for dimer formation.

The previously determined NTD tetramerisation domain spanning from 18-52 (4D2H) (Davies et al., 2015) has a 21 amino acid overlap with my NTD structure. Therefore, through alignment of the two models a structure of the N-terminal oligomerisation domain spanning residues 18-136 can be modelled. SEC-SAXS provided evidence to support this model with an envelope encompassing the modelled CtIP residue 18-136 structure. The alignment between the modelled structure and SEC-SAXS envelope provide further evidence that the dimer of dimers orientation forms a long-coiled coil rather than the C-termini of the coils in close proximity, leading to a model where there is potential for the two pairs of functionally significant C-terminal regions to be held at a distance from one another. Future work within this project area would be to confirm these predictions. Fluorescence Resonance Energy Transfer (FRET) could be used to determine the intramolecular distances between regions of the 18-145 protein and in $CtIP_{FL}$. The long term aim of this project would be to build on preliminary Cryo-EM micrographs to determine the complete structure of the CtIP N-terminal domain.

Chapter 4, CtIP C-terminal domain

4.1 Introduction

Sequence conservation locates two regions at the N- and C-termini of vertebrate CtIP that are especially important in CtIP function (Sartori et al., 2007). The C-terminal domain (CTD) comprises a stretch of ~100 amino acids located at the C-terminal region of the protein. Clear similarity to the CTD sequence is present in CtIP homologues of varying species from lower animals to plants and fungi, highlighting its significance (Andres et al., 2015; Clerici et al., 2005; Limbo et al., 2007; Sartori et al., 2007; Uanschou et al., 2007).

The functional importance of the CtIP-CTD is demonstrated by genetic conditions such as Seckel syndrome, caused by a splice mutation resulting in a C-terminal 152 amino acid truncation of the CtIP protein, (Qvist et al., 2011). *In vitro* studies using cells expressing a CtIP C-terminal 108 amino acid truncation resulted in a reduction in Chk1 phosphorylation and RPA recruitment with camptothecin-induced DSBs (Sartori et al., 2007). In the same study laser micro-irradiation showed a decrease in RPA-ssDNA formation associated with this truncated form of CtIP. CtIP interaction partners mapped to the CTD including MRN and And1 have been identified through co-immunoprecipitation experiments, using both HeLa nuclear extracts and whole-cell chromatin-free preparations respectively (Chen et al., 2017; Sartori et al., 2007). Yeast-2-Hybrid experiments have identified further CTD interaction partners such as the E3 ubiquitin ligase SIAH1. CtIP has also been shown to possess a second zinc-binding domain at the C-terminal end, binding zinc at a 1:1 ratio through two cysteine residues C813/ C816 (Sun, 2013). The function of both zinc binding events however has yet to be determined.

Within the C-terminal region are many post-translational modification sites such as the highly conserved CDK phosphorylation site threonine 847 (T847) (figure 4.1). Studies in the yeast homologue of CtIP, Sae2, have identified serine 267 as a functional equivalent of T847, highlighting the importance of the conserved phosphorylation site (Huertas et al., 2008; Huertas and Jackson, 2009a). Mutational analysis of T847 to either alanine or glutamic acid showed enhanced radiation-induced chromosome rearrangements, suggesting the importance of CDK regulation of CtIP DNA-end resection and cell cycle control (Huertas and Jackson, 2009a). It is now established that CDK phosphorylation at T847 is essential for MRN's endonuclease activity in humans (Anand et al., 2016). Conversely, the Sae2 serine 267 is dispensable for MRX activity in yeast (Cannavo and Cejka, 2014). Additional sites within the human CTD, such as the ATM/ATR phosphorylation site threonine 859 (Peterson et al., 2013; Wang et al., 2013) and the lysine 896 sumoylation site (Soria-Bretones et al., 2017) have also been reported to

be important for end-resection. Collectively these studies illustrate the functional importance of the C-terminal domain. To improve our limited understanding of CtIP's biochemical function, a detailed study of the conserved C-terminal region was performed.

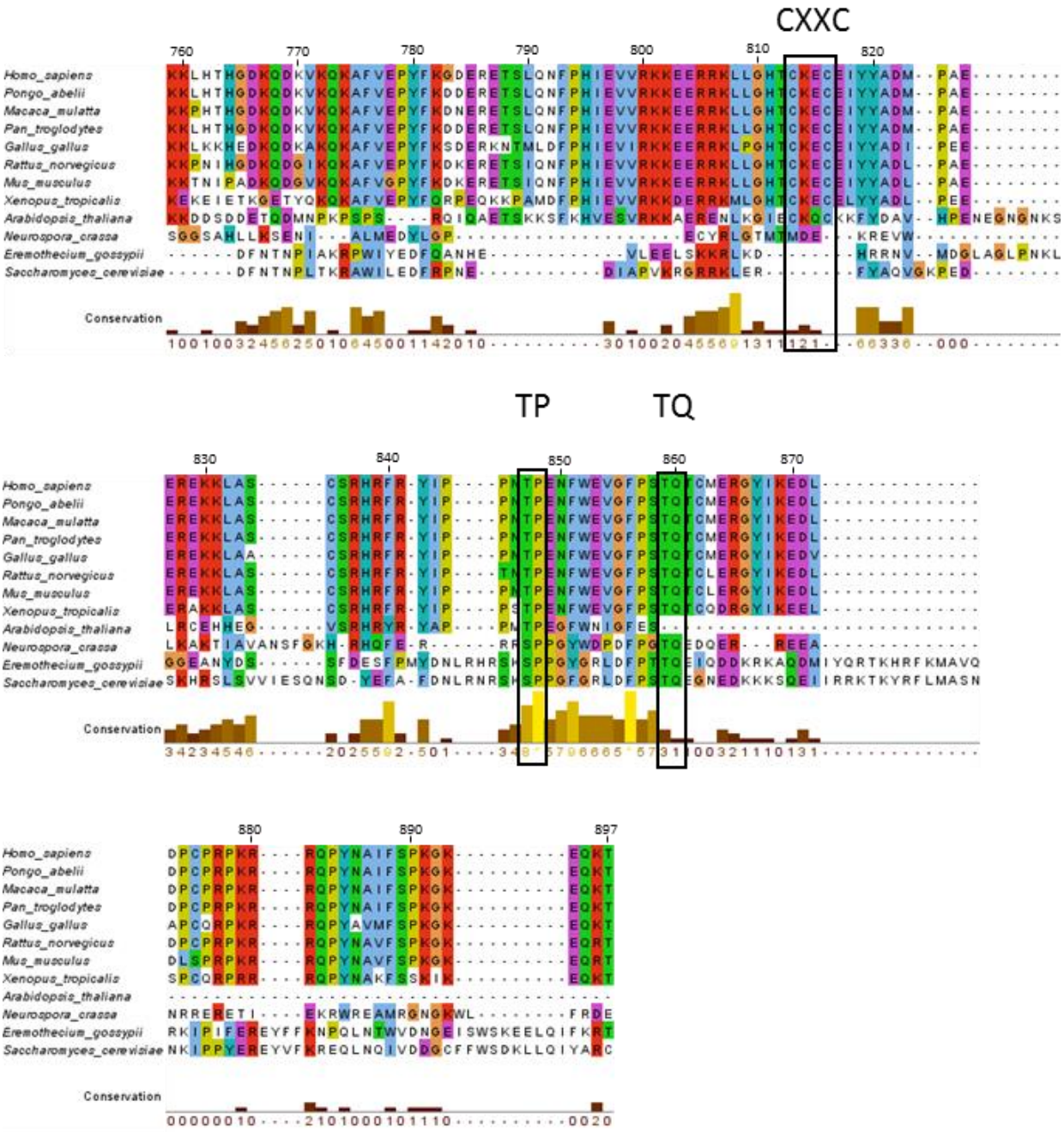


Figure 4.1: **Conservation of the CtIP C-terminal region in eukaryotes.** ClustalW (Larkin et al., 2007) sequence alignment using default settings with conservation shown. Zinc-binding residues C813/C816, TP CDK and TQ ATM/ATR sites are shown.

4.2 Results

4.2.1 DNA binding of the CtIP-CTD

For both human CtIP and the *S. pombe* homologue ctp1, DNA-binding has been localised to the C-terminal domain (Andres et al., 2015; Davies et al., 2015). To more thoroughly characterise this interaction, DNA-binding experiments were performed with purified human CtIP-CTD. The results of these experiments are described below.

4.2.1.1 Initial purification of CtIP-CTD

A CtIP-CTD construct spanning residues 769-897 was expressed and purified in accordance with the protocol developed by the former PhD student Meidai Sun (Sun, 2013). The protein was purified from Rosetta™ 2 (DE3) cells transformed with the pMAT11 CtIP-CTD construct as a His-MBP fusion protein. Using Ni-NTA and amylose chromatography with addition of reducing agent protein was purified and His-MBP tag was removed by TEV protease. Protein eluted in a single main peak on SEC with fractions containing CtIP-CTD pooled and concentrated for analysis (figure 4.2.1.1).

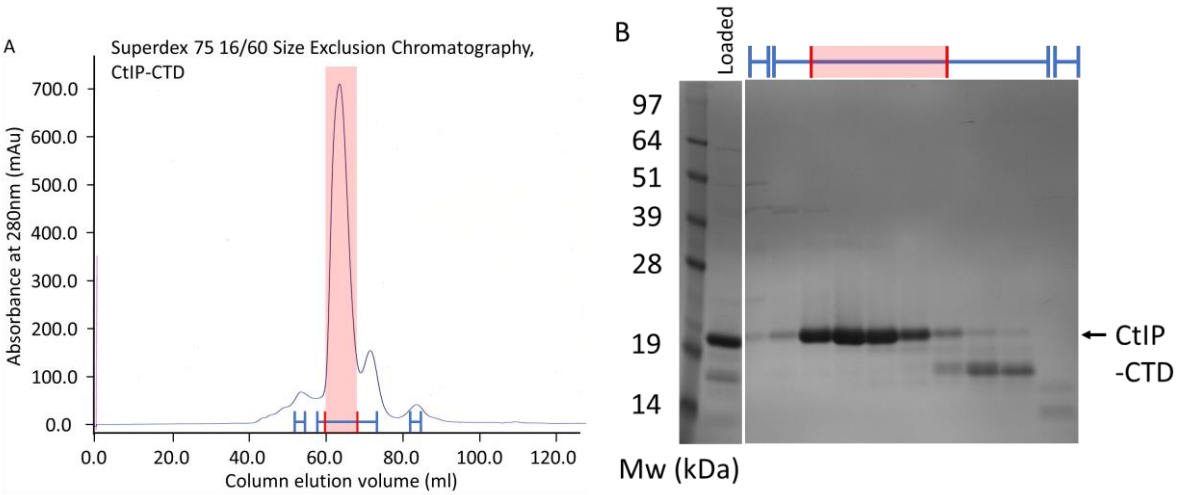


Figure 4.2.1.1: **Purification of CTD according to Sun, 2013. A)** Chromatogram of the CTD elution profile of Superdex 75 16/60 SEC. Fractions indicated were analysed by SDS-PAGE in B. **B)** SDS-PAGE analysis of SEC elution fractions. Fractions in red were pooled and concentrated for experiments.

4.2.1.2 Zinc dependency of CtIP-CTD DNA binding

The interaction of CtIP-CTD with dsDNA has previously been demonstrated using a 200 bp substrate in an EMSA (Davies et al., 2015). Using the same method, binding of a 60 bp fluorescently-labelled dsDNA substrate was also confirmed (figure 4.2.1.2b). DNA binding was further assessed by fluorescence polarisation (FP).

Using FP, formation of a CtIP-CTD – DNA complex was shown to be dependent on the addition of 50 µM $ZnAc_2$ (figure 4.2.1.2a). In contrast, zinc was not necessary for DNA binding by the same protein sample in an EMSA (figure 4.2.1.2b).

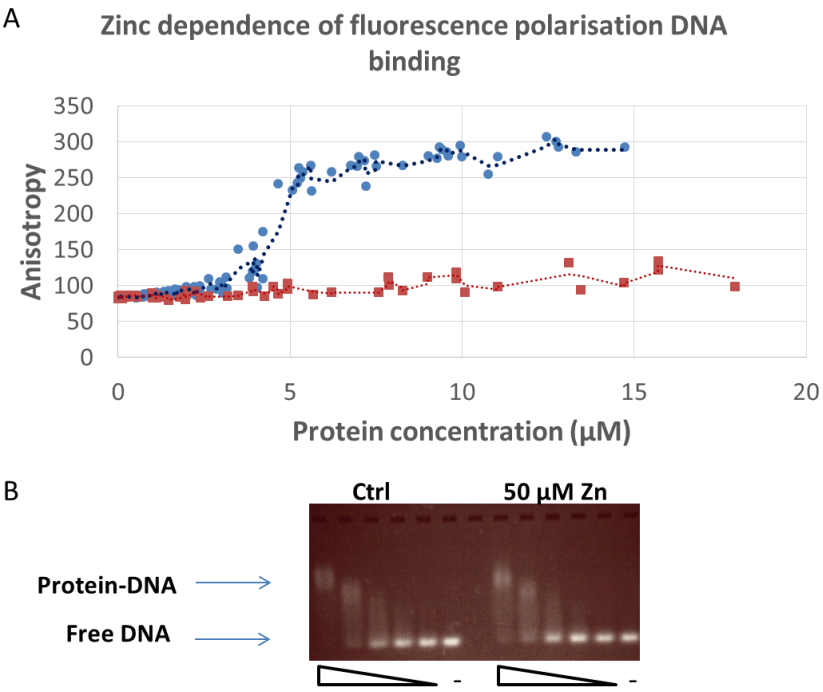


Figure 4.2.1.2: **Zinc dependency of FP DNA binding versus EMSA DNA binding.** DNA binding with linear dsDNA fdx60. **A)** Graphical representation of eight independent FP experiments with protein buffer containing 50 µM $ZnAc_2$ – blue circles. Five independent FP experiments without $ZnAc_2$ – red squares **B)** EMSA with decreasing amounts of CtIP-CTD from 10-0.62 µM with 1 µM 60 bp dsDNA. Different protein concentrations were made by serial dilution in both experiments. A moving average trend line is fitted to the

4.2.1.3 DNA binding by CTD_{DM}

Metal binding of CtIP-CTD was explored by Meidai Sun in her thesis. Two residues, C813 and C816, were determined by alanine mutagenesis to be responsible for zinc binding (Sun, 2013). Therefore, if DNA binding is dependent on these residues for zinc incorporation, mutation to alanine (figure 4.2.1.3-1) should inhibit DNA binding.

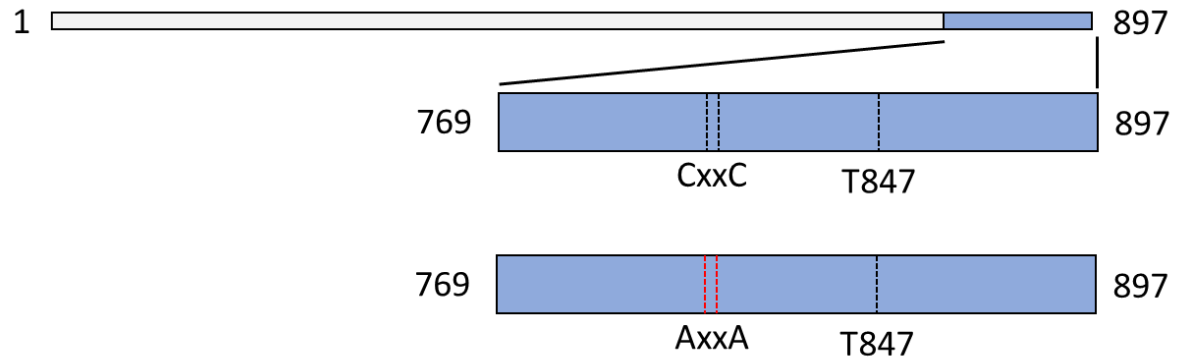


Figure 4.2.1.3-1: **Schematic drawing of CtIP-CTD and CTD_{DM}.** CtIP residues 769-897 are shown in blue, with zinc binding site (CXXC) and T847 CDK site. For CTD_{DM} the zinc-binding site is mutated to alanine (AXXA).

Using the same purification procedure, a CTD C813A/C816A (CTD_{DM}) protein was purified to homogeneity (figure 4.2.1.3-2); and incorporation of the two mutations was confirmed by mass-spectrometry (appendix supplementary figure 7.4).

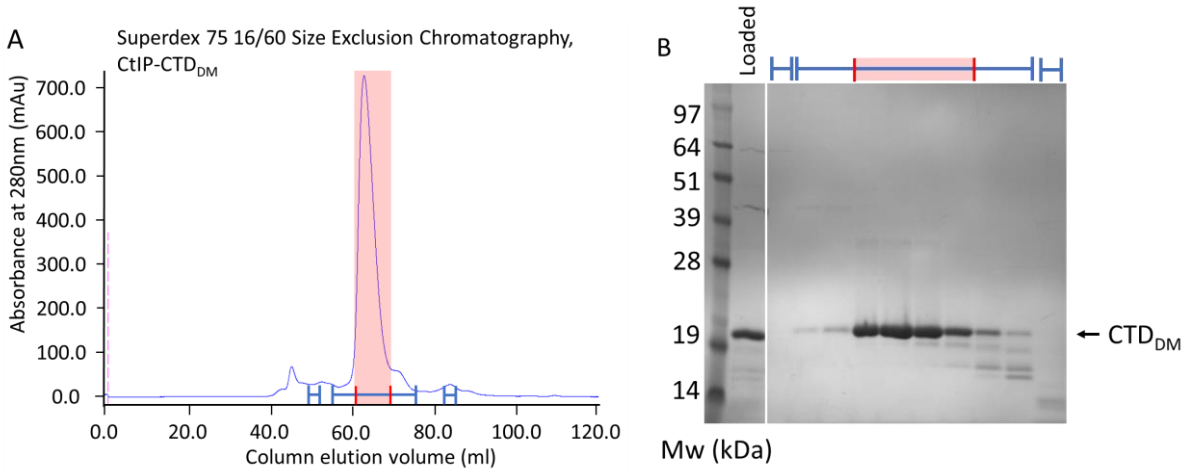


Figure 4.2.1.3-2: **Purification of CtIP-CTD_{DM} according to Sun, 2013. A)** Chromatogram of the CTD elution profile of Superdex 75 16/60 SEC. Fractions indicated were analysed by SDS-PAGE in B. Elution peak is at 64 mL **B)** SDS-PAGE analysis of SEC elution fractions. Fractions in red were pooled and concentrated for experiments.

DNA-binding experiments by EMSA were performed in triplicate with no differences observed between the mutated and non-mutated protein (figure 4.2.1.3-3). Preliminary FP studies using CTD_{DM} have also shown similar DNA binding compared to CtIP-CTD; zinc dependency was also still apparent for CTD_{DM} DNA binding by FP (figure 4.2.1.3-4). These results indicate that the two residues identified for zinc binding in CTD_{DM} are not responsible for the zinc dependency for DNA binding by FP and requires further zinc binding analysis.

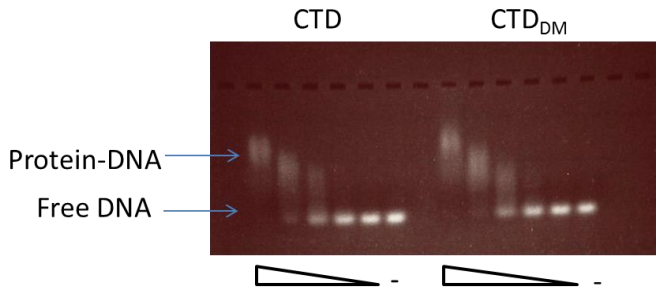


Figure 4.2.1.3-3: **CTD_{DM} binds DNA by EMSA.** EMSA with decreasing amounts of protein from 10-0.62 µM with 1 µM 60bp dsDNA, resulting in the formation of slow migrating species which represent protein-DNA complexes. Different protein concentrations were made by serial dilution with identical concentrations for CTD (left hand side) and CTD_{DM} (right hand side).

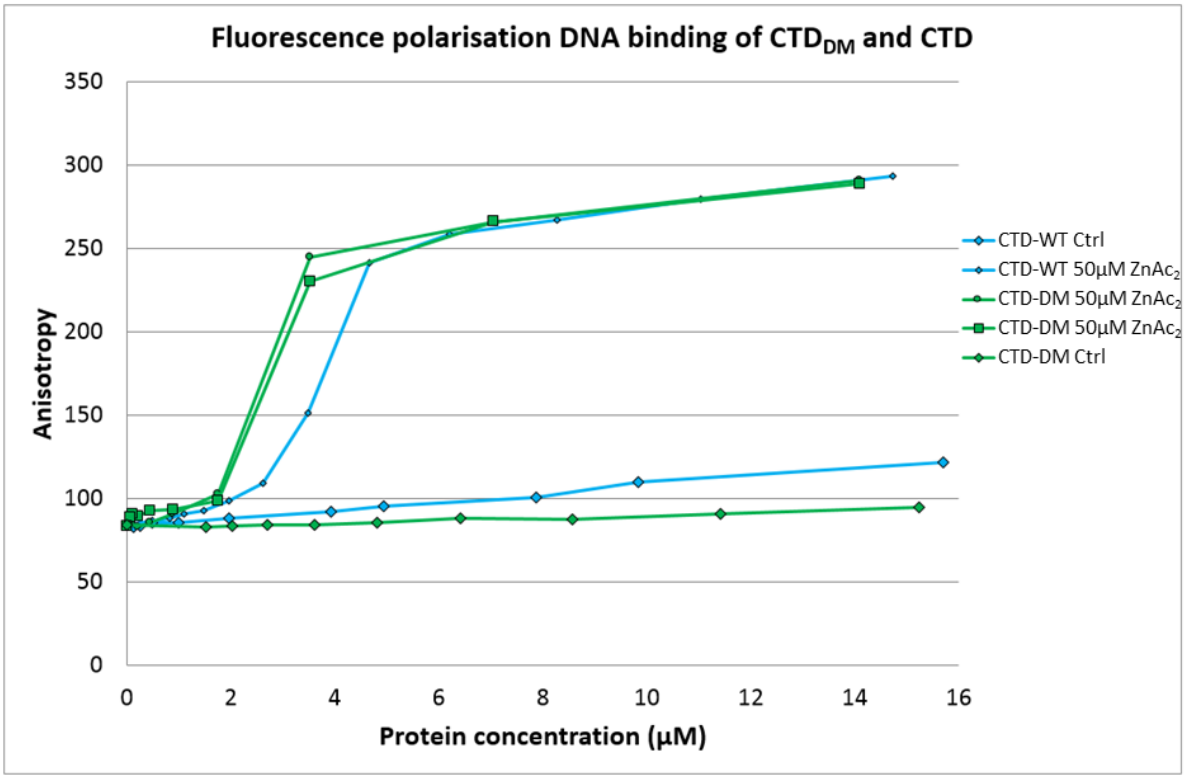


Figure 4.2.1.3-4**: Zinc dependency of FP DNA binding for CTD_{DM}.** DNA binding with linear dsDNA fdx60. Graphical representation of two CTD_{DM} binding curves with buffer containing 50 µM $ZnAc_2$ (green squares) or without zinc (green diamonds). Blue traces represent CtIP-CTD binding curves with (circles) and without (diamonds) 50 µM $ZnAc_2$.

4.2.1.4 Purification optimisation to enhance zinc incorporation

The original purification for CtIP-CTD included a nickel affinity chromatography step, and it was shown that this resulted in nickel incorporation into the recombinant CtIP-CTD (Sun, 2013). In the light of the observed zinc dependency of DNA binding by FP, the purification protocol was adapted to ensure zinc incorporation into the CtIP-CTD over other metals. This would allow quantitative analysis of zinc incorporation and its effect on DNA binding.

The nickel-affinity purification step was therefore eliminated from the protocol. The CtIP-CTD was initially purified from clarified lysate by amylose affinity chromatography, utilising the MBP tag of the fusion protein prior to TEV cleavage. Three procedures were tested for removal of contaminants. Zinc-charged NTA resin bound CtIP-CTD without removing contamination (figure 4.2.1.4-1). Optimisation of a heparin chromatography stage and a second amylose step after TEV cleavage enabled removal of most contaminants (figure 4.2.1.4-1).

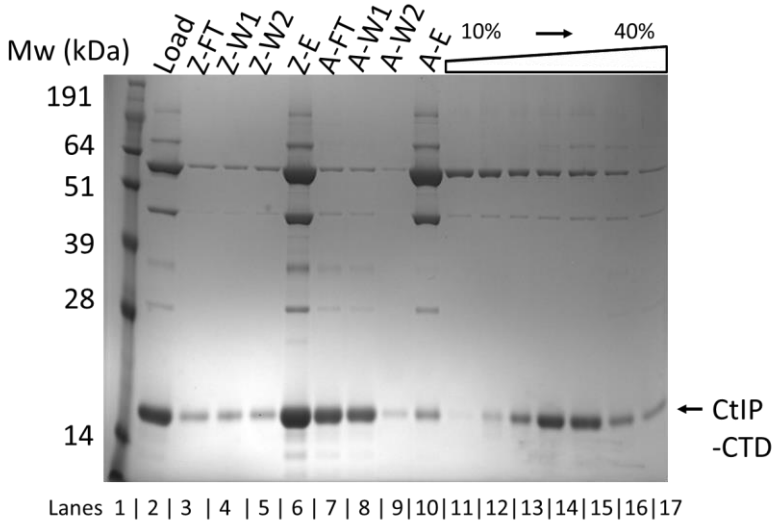


Figure 4.2.1.4-1: **SDS-PAGE analysis of CtIP-CTD purification optimisations.** Position of CtIP-CTD is indicated on the gel. From left to right: Protein applied to onto each purification column (load). "Z" stands for Zn-NTA, "A" for Amylose, "FT" for flow-through, "W" for wash and "E" for elution. **Lanes 3-6** show recapture using a Zn-NTA column with most of the protein in the elution fraction. **Lanes 7-10** show recapture using an amylose column with most of CtIP-CTD within the flow-through and wash fractions while MBP-CTD and MBP remain bound. **Lanes 11-17** are fractions from a 5 mL HiTrap Heparin column elution from 10-40% v/v buffer B (1M NaCL) in 5% v/v buffer B steps. Overlapping peaks were present with most of the contaminants eluted between 15-20% v/v B; 17.5% v/v B was used for further purifications.

Further optimisations to the protocol included the addition of final concentration 250 µM zinc to the growth media to enhance incorporation. This decreased the proportion of higher molecular weight species on Superdex 75 16/60 SEC (figure 4.2.1.4-2, P1) and increased the yield of CtIP-CTD for the

same starting volume of cells (figure 4.2.1.4-2, p3). An earlier eluting peak (figure 4.2.1.4-2, P2) persisted irrespective of purification procedure at a larger weight than CtIP-CTD.

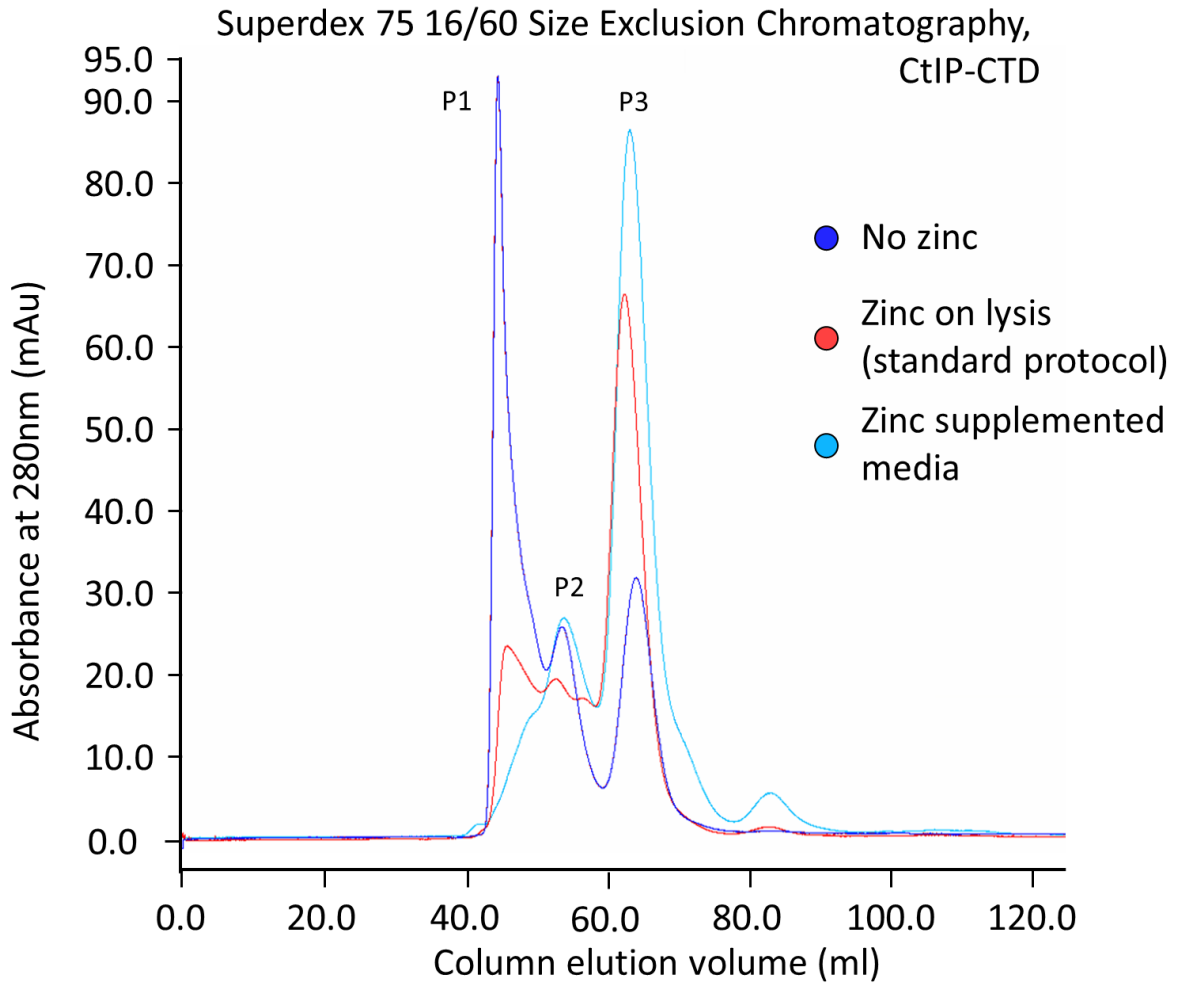


Figure 4.2.1.4-2: **SEC traces for CtIP-CTD prepared according to three different zinc addition protocols**. Blue trace: no zinc in the buffer or the growth media. Red trace: zinc added on cell lysis. Cyan trace: zinc supplemented media without further addition of zinc.

A new purification strategy evolved from the series of optimisations. The CtIP-CTD construct was expressed in two litres of Rosetta™ 2 (DE3) cells with or without zinc supplemented media. Following lysis and clarification, supernatant was purified by amylose chromatography with elution before overnight dialysis with TEV cleavage. The TEV protease was pre-incubated with DTT with no DTT added to the dialysis buffer. Hi-Trap heparin chromatography was used to remove contaminants using a 17.5% v/v buffer B (1 M NaCl) step with a gradient elution. Where necessary fractions containing CtIP-CTD were further purified by a final amylose chromatography stage either before or after concentration to 1 ml for application to the Superdex 75 16/60 SEC chromatography. Fractions containing CtIP-CTD were analysed throughout the process by SDS-PAGE (figure 4.2.1.4-3).

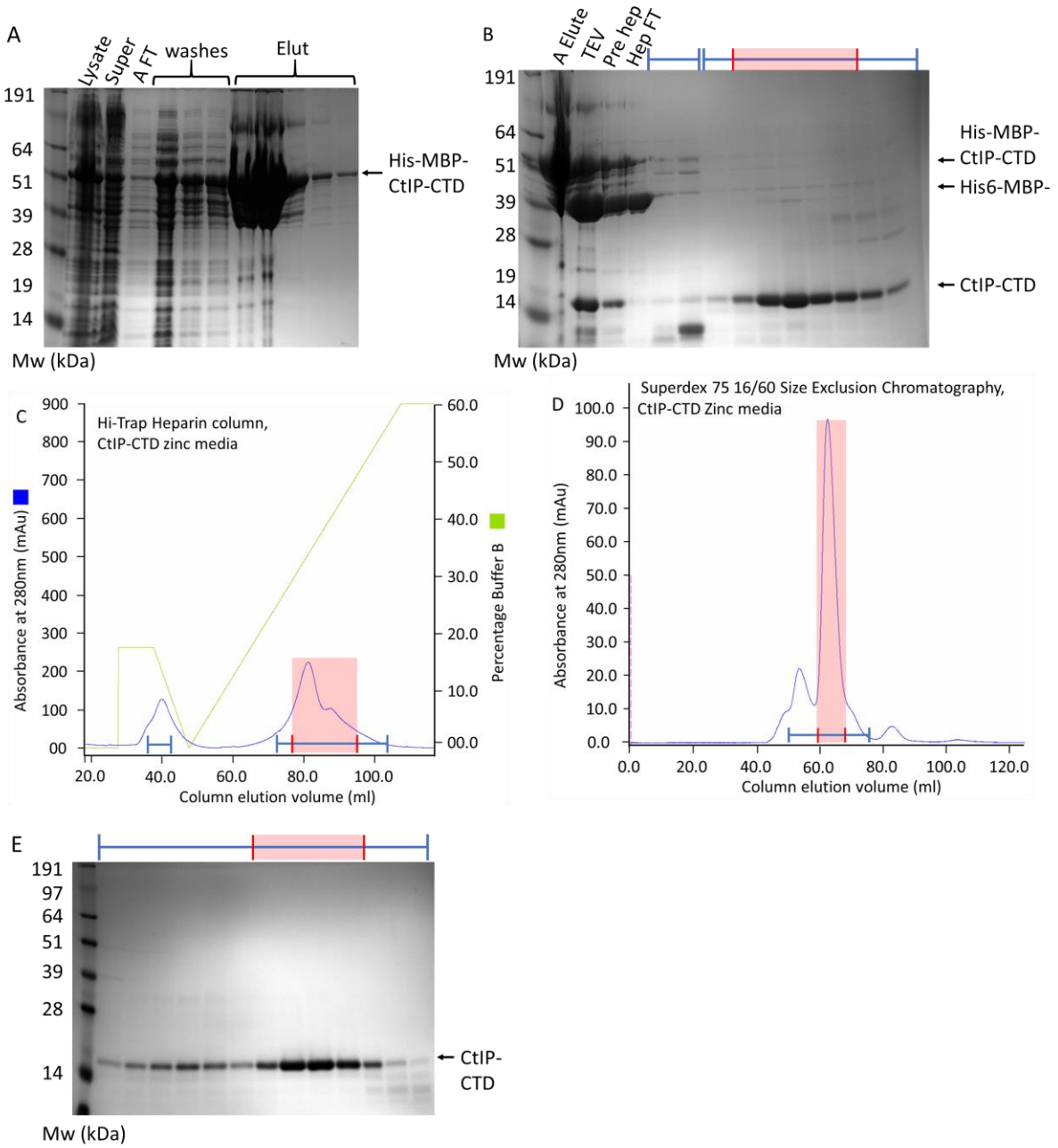


Figure 4.2.1.4-3: **Purification of CtIP-CTD by amylose and heparin chromatography**. CtIP-CTD was expressed using zinc supplemented media **A)** SDS-PAGE analysis of CtIP-CTD purification by Amylose affinity chromatography (A), with flow-through (FT), washes and 10 mM maltose elution fractions (Elut) shown. **B)** Elution fractions 1-3 were pooled for TEV cleavage (TEV), followed by salt dilution (Pre-Hep) any application to a heparin column. Fractions indicated in red/ blue correspond to C. **C)** Hi-Trap heparin chromatography chromatogram, buffer B was 1 M NaCl with a 17.5% v/v step removing contaminants. Fractions in red were pooled for further processing. For this purification fractions pooled were concentrated to 1 mL for Superdex75 16-60 SEC in D. **D)** Chromatographic profile of CtIP-CTD elution, fractions indicated were analysed by SDS-PAGE in E. **E)** Fractions in red were pooled as CtIP-CTD and flash frozen in liquid nitrogen.

4.2.1.6 DNA binding of CtIP-CTD purified from zinc supplemented 2xYT media

The DNA-binding ability of the CtIP-CTD expressed in zinc supplemented media was tested by FP and EMSA to determine if the modified purification method had removed the previously observed necessity for additional zinc in the FP experiments (figure 4.2.1.2). Preliminary FP results showed that the zinc dependency persisted, whereas DNA binding by EMSA consistently showed no difference (figure 4.2.1.6-1). The requirement for the presence of zinc in the buffer used in the FP experiments was not investigated further, and DNA binding by the CtIP-CTD was studied using the EMSA using the three different additions of zinc. Zinc binding by the CtIP-CTD was further investigated in section 4.2.2.

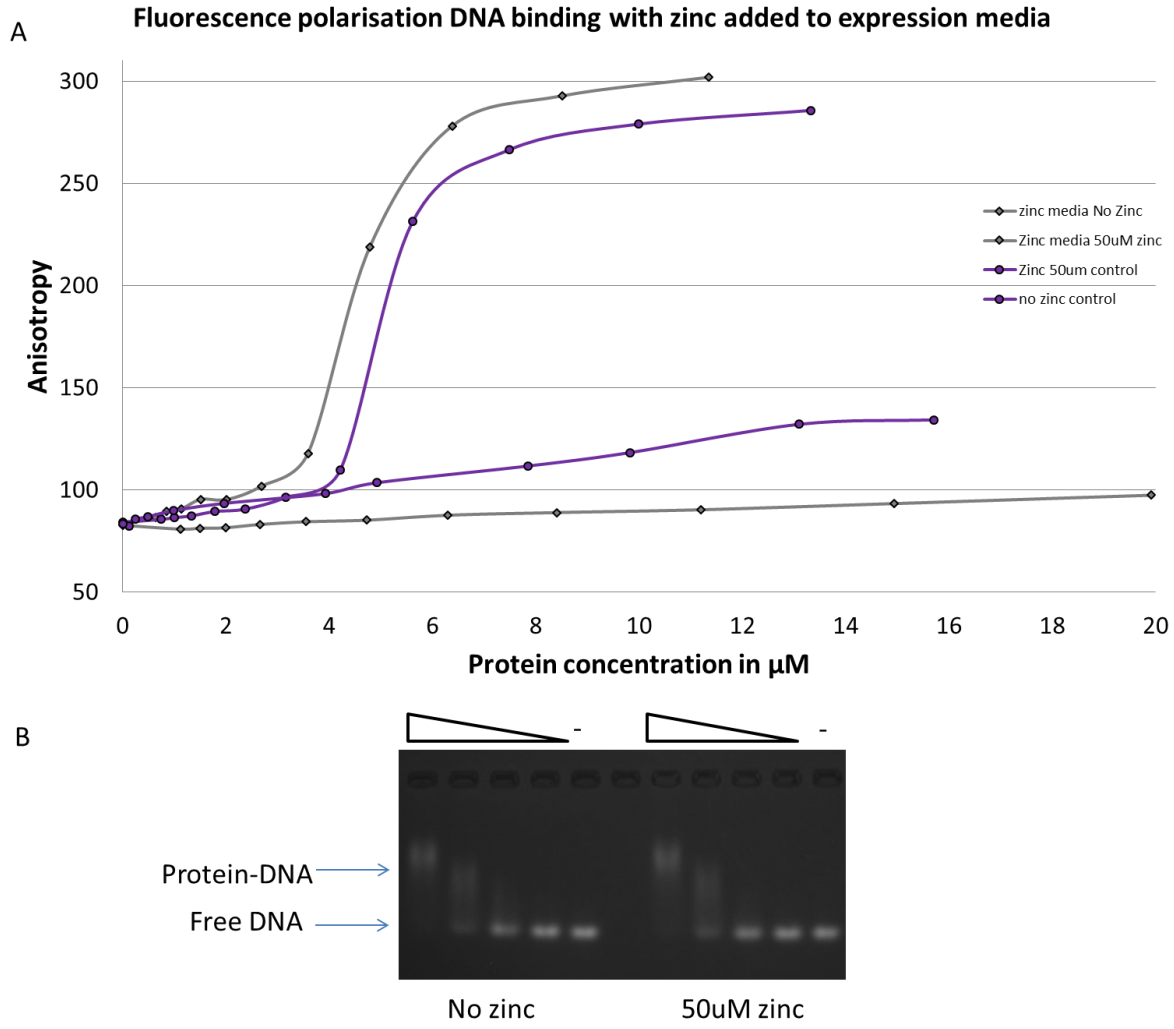


Figure 4.2.1.6-1: **Zinc dependency of FP DNA binding for CtIP-CTD expressed using zinc supplemented media verses EMSA DNA binding**. DNA binding with linear dsDNA fdx60. **A)** Graphical representation of a single CtIP-CTD binding curve: Grey represents zinc media with or without additional 50 µM zinc. Purple represents the previous purification method with or without additional 50 µM zinc. **B)** EMSA with decreasing amounts of CtIP-CTD grown in the presence of zinc supplemented media from 12-1.5 µM, resulting in the formation of slow migrating species which represent protein-DNA complexes. EMSA was repeated three times with a representative gel shown. Different protein concentrations were made by serial dilution in both experiments. Were 50µM zinc is indicated this is 50µM $ZnAc_2$.

Optimisations to the purification strategy did not alter the ability of CtIP-CTD to bind DNA, as analysed by EMSA. Overall, both CtIP-CTD and CtIP-CTD_{DM} were observed to display similar DNA binding characteristics that were broadly unchanged by the purification method chosen (figure 4.2.1.6-2). While replicates are yet to be performed, these will allow conclusive analysis of the small differences seen between CTD_{DM} purification methods.

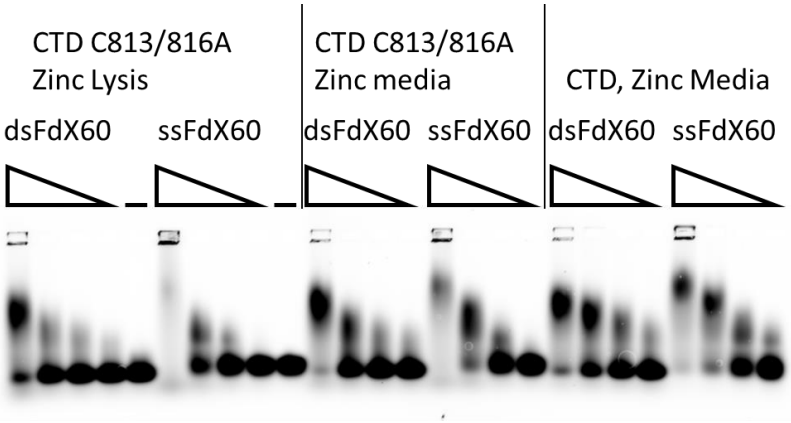


Figure 4.2.1.6-2: **CtIP-CTD_{DM} binds DNA irrespective of purification method as shown by EMSA.** An EMSA with decreasing amounts of CTD from 10-1.5 µM, resulting in the formation of slow migrating species which represent protein-DNA complexes. LTR, CTD_{DM} zinc on lysis, CTD_{DM} zinc media, CTD zinc media each with 100 nM 60 bp fluorescent dsDNA or ssDNA as indicated.

4.2.1.7 Minimum DNA length required for CtIP-CTD interaction

EMSA results were consistent for CtIP-CTD showing binding to a 60 bp dsDNA oligonucleotide. To deduce the minimum length of DNA necessary for the CtIP-CTD to DNA interaction further EMSAs were performed using the same procedure with progressively shorter dsDNA. Initial experiments showed DNA binding to 40 bp dsDNA with minimal signs of DNA binding for substrates of 20 bp dsDNA and below (figure 4.2.1.7). The length of substrates was reduced by 5 bp dsDNA iteratively, with similar amounts of binding shown from 40 bp to 25 bp dsDNA (figure 4.2.1.7). The DNA binding of CtIP to 20 bp is greatly impaired with no binding at low concentrations and a large reduction at 5.0 µM. These results indicate the minimum length requirement of DNA for CtIP binding to be between 25 bp and 20 bp.

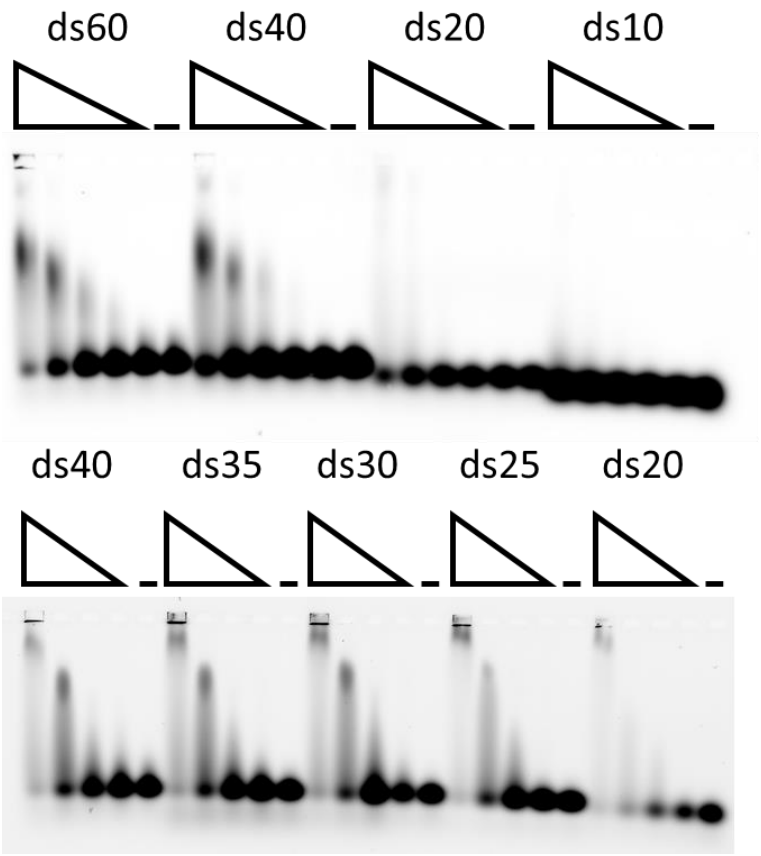


Figure 4.2.1.7**: Determination of minimum DNA length for CtIP-CTD: DNA interaction.** Top panel: triangle represents decreasing concentrations of CtIP-CTD from 5.0 µM down to 0.31 µM. LTR EMSAs with 60, 40, 20 and 10 bp fluorescently-labelled dsDNA at 100 nM. Bottom panel – triangle represents decreasing CtIP concentrations from 5.0 µM to 0.62 µM. LTR, 40, 35, 30, 25 and 20 bp fluorescently labelled dsDNA at 100 nM. Addition of CtIP-CTD results in the formation of a slower migrating species which represents protein-DNA complexes. Protein was prepared by serial dilution, with both experiments repeated in triplicate.

4.2.2 Zinc incorporation by the C-terminal domain

Reminiscent of the NTD, the CTD of CtIP has been shown to contain a zinc-binding domain. Previously, two cysteine residues (amino acids 813 and 816) were shown to be necessary for zinc chelation (Sun, 2013). However, results described above are at odds with this observation, as the CTD_{DM} appeared to be still able to bind DNA by FP in the presence of 50 µM zinc. To address these discrepancies, the zinc content of the CtIP-CTD was analysed using spectrophotometry analysis. The chosen spectroscopic assay involves the use of the metal chelator 4-(2-pyridylazo)-resorcinol (PAR), which undergoes a wavelength shift in peak absorbance upon binding to zinc. This allows quantification of zinc content in the sample by comparison to standards of known zinc concentrations.

The binding of PAR is not specific for zinc, but extends to nickel, copper, cobalt and other divalent metals. This method has therefore been established for many years to determine trace levels of different metalloproteins within a sample (Ahrland and Herman, 1975; Hunt et al., 1985; Jezorek and Freiser, 1979; Säbel et al., 2009). The spectrophotometric profile of PAR bound to copper and cobalt is different to that in the presence of nickel and zinc, allowing discrimination between these two sets of metals; in contrast, the spectra of PAR-bound nickel and zinc are indistinguishable from each other for an unknown concentration of metal within a sample (Säbel et al., 2009). Therefore, to determine

zinc concentration reliably using the PAR assay, precautions must be taken to avoid exposure of the protein sample to nickel.

Analysis of the PAR absorption spectrum over the range of 300 to 600 nm (figure 4.2.2-1), shows that free PAR has a peak at 414 nm. With addition of increasing zinc acetate concentrations, a second peak at 492 nm gradually forms representing the PAR-Zn^{2+} complex, whereas the 414 nm free PAR peak decreases. An example of PAR's absorption profile in the presence of increasing concentration of $ZnAc_2$ is shown in figure 4.2.2-1. For each measurement of zinc content, a new standard curve was generated.

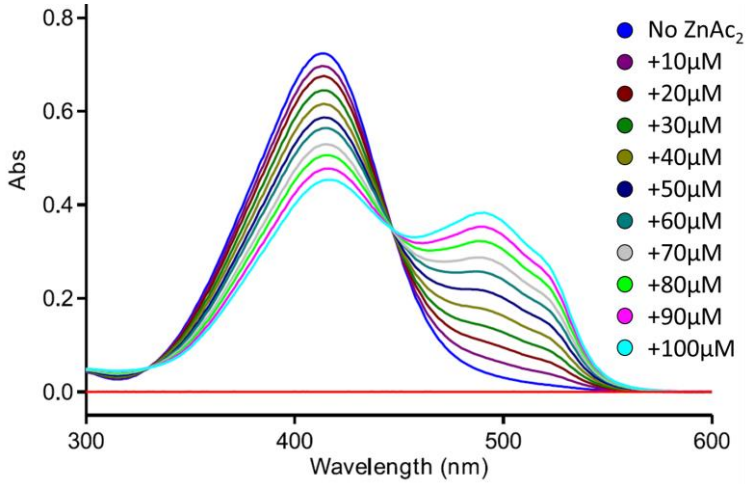


Figure 4.2.2-1: **Spectrophotometric analysis of zinc standard concentrations in complex with PAR**. Concentration range of $ZnAc_2$ from 0 (blue) to 100 nM (cyan) $ZnAc_2$.

Using the original Ni-NTA purification method, the PAR assay yielded a CtIP-CTD to Zn^{2+} ratio of 1:1 over 3 repeats, similar to that reported previously (Sun, 2013). Using the amylose purification procedure that excluded exposure to Ni^{2+} using the same zinc addition on lysis as the original purification method a CtIP-CTD to Zn^{2+} ratio of 2.2:1 over 3 independent repeats was measured, which is lower than previously reported (figure 4.2.2-2). A possible explanation for the difference is a contribution from nickel ions bound to the protein that had been purified using the original method.

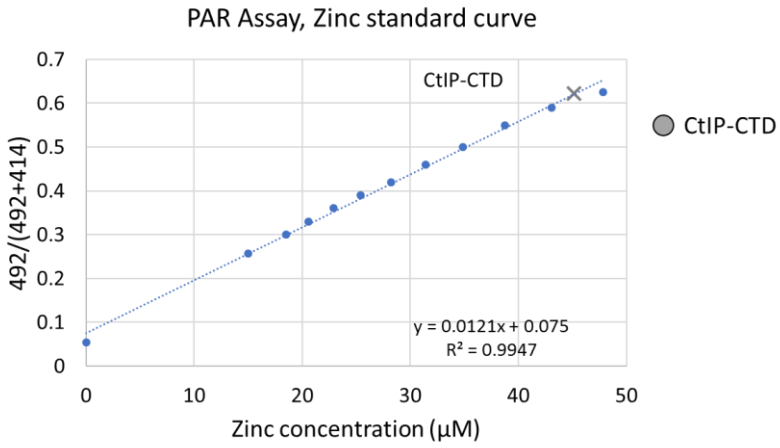


Figure 4.2.2-2: **Determination of zinc content for CtIP-CTD.** Blue circles represent values used in the linear fit with equation displayed. Grey Cross plotted on the line is the average ratio of absorbance values at 492 and 414 nm for the CtIP-CTD sample (100 µM) plotted on the linear fit.

Using the new purification procedure, and in contrast to previous results by Meidai Sun (Sun, 2013), concentrations of zinc recovered from CTD_{DM} were also at a ratio of 2.4:1 protein to zinc (figure 4.2.2-.3).

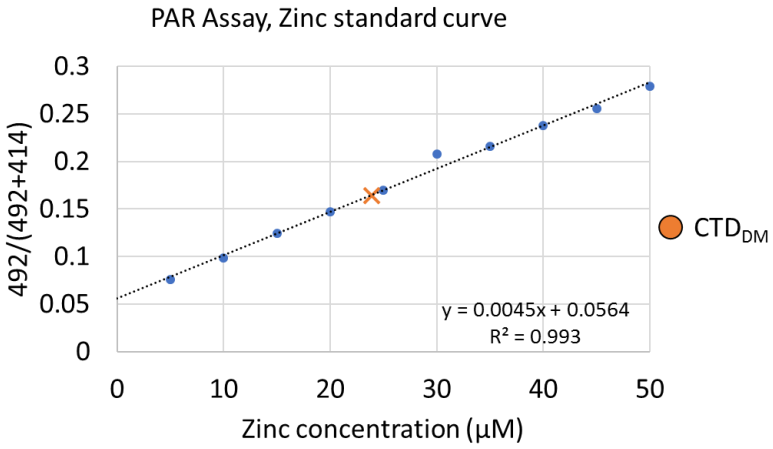
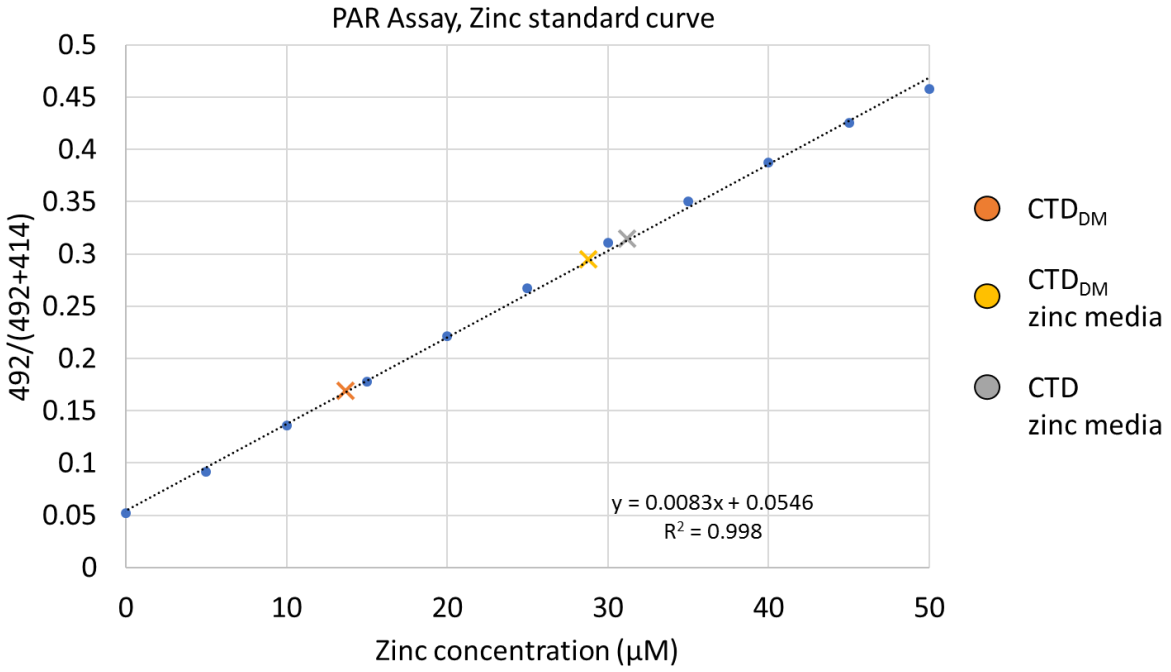


Figure 4.2.2-3: **Determination of zinc content for CTD_{DM}.** Blue circles represent values used in the linear fit with equation displayed. Orange cross on the line is the average ratio of absorbance values at 492 and 414 nm for the CTD_{DM} sample (57 µM).

The conclusion from these experiments is that removal of Ni-NTA chromatography from the purification protocol has led to a reduction in the total metal content of CtIP-CTD preparations, suggesting that part of the metal previously detected was due to nickel contamination. It was further observed that DNA binding by CtIP-CTD, when analysed by FP, was dependent on the addition of zinc to the sample, suggesting that the zinc content in original CtIP-CTD preparations might be sub-optimal. Together, these observations led to exploration of additional strategies to improve zinc incorporation in the CtIP-CTD.

The addition of zinc to bacterial growth media is a well-documented method for the soluble over-expression of exogenous zinc-binding proteins. A typical approach is to add 250 µM zinc acetate to the growth media prior to induction (Graham et al., 2009; Lee et al., 2014; Yamamoto and Ishihama, 2005). Using zinc-supplemented media increased the levels of Zn^{2+} ions recovered from CtIP-CTD to the level of 1:1 protein to zinc. Surprisingly, CTD_{DM} also showed a similar level of Zn^{2+} ion incorporation, with the addition of zinc to the bacterial media giving a protein to zinc ratio of 1:1 over three repeats (figure 4.2.2-4). The level of zinc recovered from CtIP-CTD and CTD_{DM} was the same 1:1 ratio as the original purification, however greater than purifying without Ni-NTA chromatography. This further supports the presence of contaminant Ni^{2+} ions bound to the CTD and suggests the C813/C816 residues mutated to alanine in CTD_{DM} were required for binding these Ni^{2+} ions.



Protein	Protein Concentration / µM	Average Zinc concentration/ µM	Ratio of Protein to zinc
CTD_{DM}	28.8	13.7 ± 0.2	2:1
CTD_{DM} – Amylose purification	32.6	31.2 ± 0.7	1:1
CtIP-CTD – Amylose purification	31.8	29 ± 2	1.1:1

Figure 4.2.2-4: **Determination of zinc content for CtIP-CTD and CTD_{DM} expressed using zinc supplemented media.** Blue circles represent values used in the linear fit with equation displayed. Orange cross on the line is the average ratio of absorbance values at 492 to 414 nm for the CTD_{DM} sample. Yellow cross on the line is the average ratio of absorbance values at 492 and 414 nm for CTD_{DM} with zinc media, grey cross on the line is the average for CtIP-CTD with zinc media. Concentrations of protein used, and their respective zinc concentrations are shown in the bottom panel.

4.2.2.1 Determining the zinc binding residues

As CTD_{DM} retained the capability to bind zinc at a 1:1 ratio other residues must also be involved and were further explored. A total of 6 cysteine or histidine residues (excluding C813A/C816A) with potential for zinc binding were selected within the CtIP-CTD by sequence analysis (figure 4.2.2.1) with individual point mutations created by overlap extension PCR.

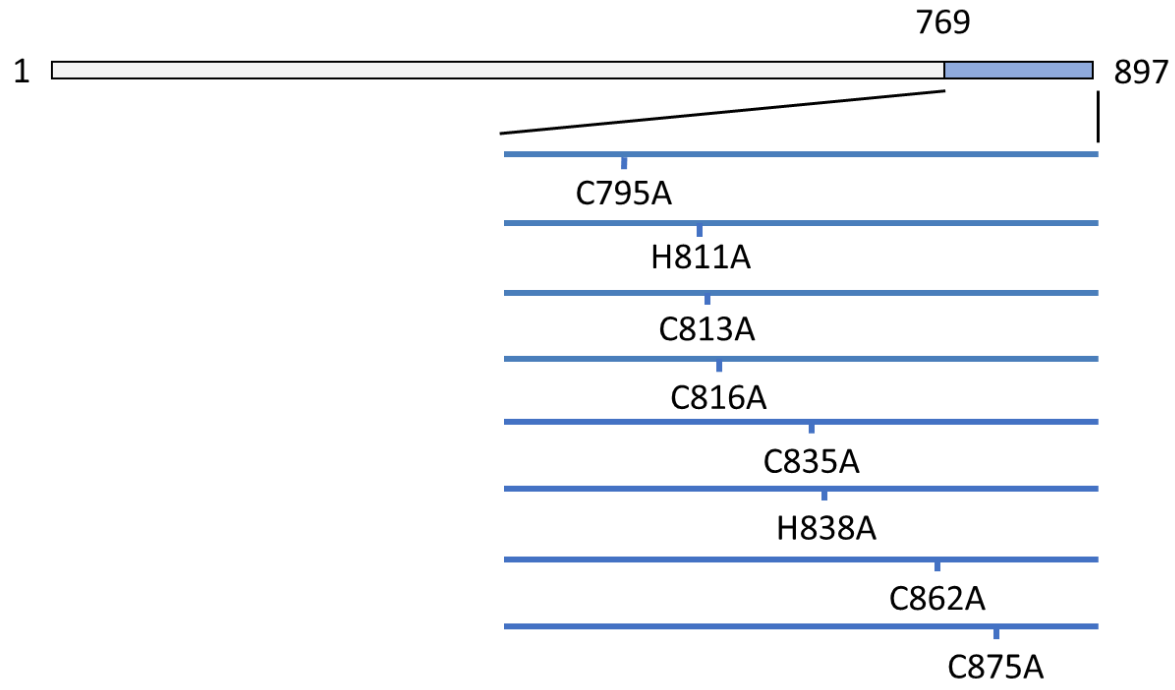


Figure 4.2.2.1: **Schematic drawing of CtIP-CTD zinc binding single point mutations.**

4.2.2.1.1 Expression and purification of CtIP-CTD point mutants

For purification of each construct two litres of cells were grown without zinc-supplemented media. Using the same protocol as for CtIP-CTD (figure 4.2.1.4-3) protein from clarified lysate was purified through sequential steps of: amylose initial capture, overnight dialysis with TEV cleavage and heparin chromatography with a 17.5% buffer B step for contamination removal. At this point the purification diverged to further remove contamination: protein was concentrated to 1 mL prior to the final amylose recapture. The flow-through from this final recapture was applied to a SEC 16/60 Superdex 75 column and resulting fractions containing CtIP-CTD were identified by SDS-PAGE. The same process was used for all six individual point mutants. Figure 4.2.2.1.1 illustrates the final purification stage for CTD_{C835A} as a representative example.

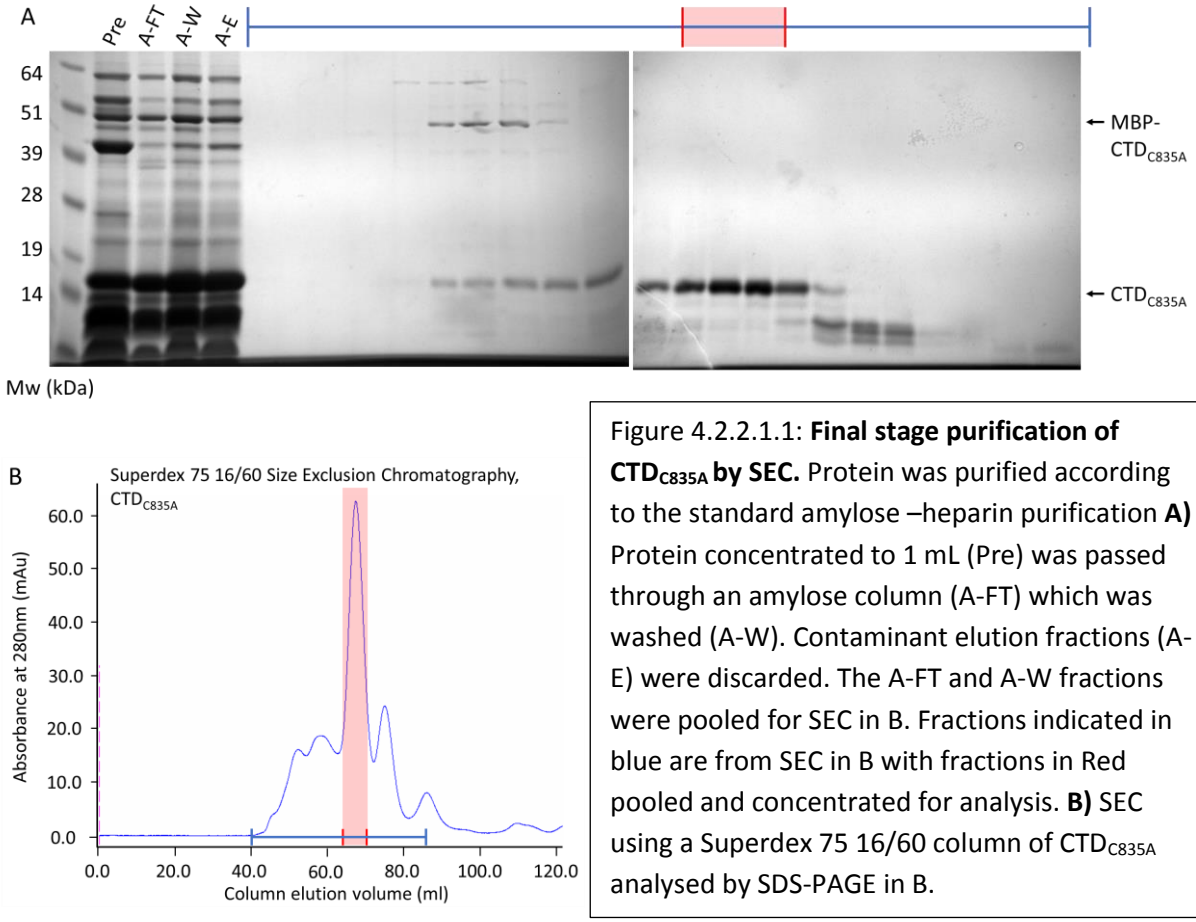


Figure 4.2.2.1.1: **Final stage purification of CTD_{C835A} by SEC.** Protein was purified according to the standard amylose –heparin purification **A)** Protein concentrated to 1 mL (Pre) was passed through an amylose column (A-FT) which was washed (A-W). Contaminant elution fractions (A-E) were discarded. The A-FT and A-W fractions were pooled for SEC in B. Fractions indicated in blue are from SEC in B with fractions in Red pooled and concentrated for analysis. **B)** SEC using a Superdex 75 16/60 column of CTD_{C835A} analysed by SDS-PAGE in B.

4.2.2.1.2 Determination of zinc content for CtIP-CTD point mutants

Using the PAR assay, as described earlier in this section, the zinc content of each individual CtIP-CTD protein mutant was analysed. Protein concentration varied from purification to purification with concentrations used denoted in table 4.2.2.1.2-1. As non-zinc supplemented media was used, the ratios were expected to be 2:1 protein to zinc.

Protein	Protein concentration used/ µM	Average zinc concentration/ µM	Protein to zinc ratio
CTD_{C795A}	80.0	40 ± 1	2.0 ± 0.5 :1
CTD_{H811A}	95.3	31 ± 1	3.0 ± 0.1 :1
CTD_{C835A}	84.7	33 ± 2	2.5 ± 0.1 :1
CTD_{H838A}	76.8	44.4 ± 0.2	1.7 ± 0.01 :1
CTD_{C862A}	84.0	44.0 ± 0.7	1.9 ± 0.1 :1
CTD_{C875A}	59.0	31 ± 4	1.9 ± 0.6 :1

Table 4.2.2.1.2-1: **Protein and zinc concentrations for CtIP-CTD mutants.** Zinc concentrations were determined through the standard curve in figure 4.2.2.1.2-2.

Zinc was detected for all 6 single point mutations through observation of a colour change on mixing proteinase K-digested protein with PAR. Slight differences can be seen within the ratio of protein to zinc compared between the mutants, with CTD_{H811A} having the least zinc present (table 4.2.2.1.2-1 & figure 4.2.2.1.2-2). Further analysis will be necessary over multiple purifications to determine any true differences between the mutations with zinc supplemented media. However, no single mutation tested abolished zinc binding.

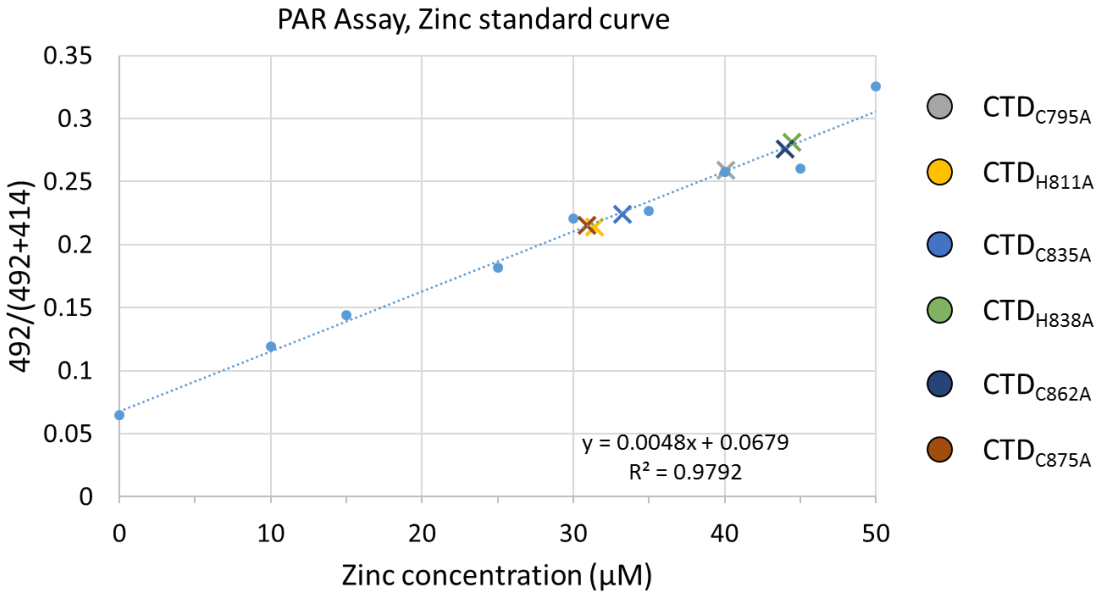


Figure 4.2.2.1.2-2: **Determination of zinc content for CtIP-CTD single zinc binding point mutations.** Blue circles represent values used in the linear fit with equation displayed. Values for 5 and 1 5µM were omitted as they did not start at the baseline. Crosses represent the average ratio of absorbance values at 492 and 414 nm for each protein.

4.2.2.2 Expression and purification of CTD_{ABC}

Single point mutations failed to abolish zinc binding of the CtIP-CTD. As there is potential for compensation between the different residues, combinations of mutations were generated by overlap extension PCR. Although many of these were cloned, due to insufficient time only one of these has been expressed. This clone, named CTD_{ABC} combines all 6 mutated sites (figure 4.2.2.2-1).

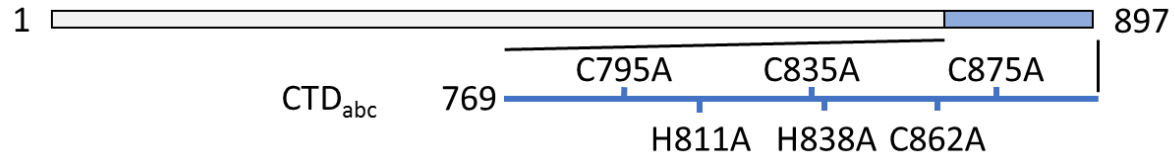


Figure 4.2.2.2-1: **Schematic drawing of CTD_{ABC} zinc binding 6 combined point mutations**

The 6 combined point mutations resulted in a large amount of degradation. The standard purification process was therefore halted prior to TEV cleavage, and the sample was subjected to sequential steps of Hi-Trap heparin chromatography and SEC (figure 4.2.2.2-2). In both steps, contaminating degradation products were observed to co-elute with the target of interest. Due to the inability to satisfactorily purify this mutant protein, analysis of zinc content was not undertaken. Further analysis of combination mutations is therefore necessary, perhaps with a change of purification approach.

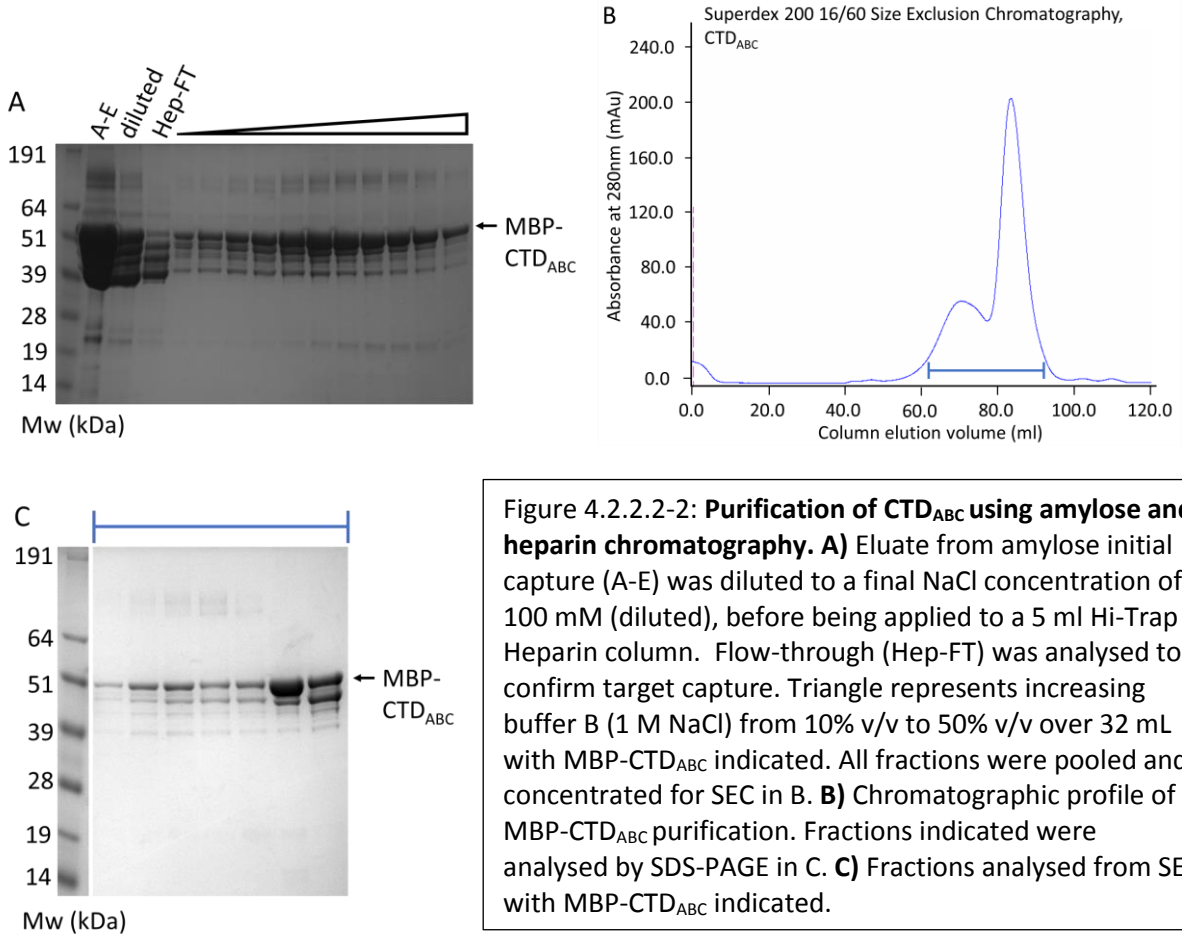


Figure 4.2.2.2-2: **Purification of CTD_{ABC} using amylose and heparin chromatography. A)** Eluate from amylose initial capture (A-E) was diluted to a final NaCl concentration of 100 mM (diluted), before being applied to a 5 ml Hi-Trap Heparin column. Flow-through (Hep-FT) was analysed to confirm target capture. Triangle represents increasing buffer B (1 M NaCl) from 10% v/v to 50% v/v over 32 mL with MBP-CTD_{ABC} indicated. All fractions were pooled and concentrated for SEC in B. **B)** Chromatographic profile of MBP-CTD_{ABC} purification. Fractions indicated were analysed by SDS-PAGE in C. **C)** Fractions analysed from SEC with MBP-CTD_{ABC} indicated.

4.2.3 Confirmation that CtIP-CTD can form dimers and their analysis.

During successive purification runs of CtIP-CTD a higher molecular weight peak/shoulder was consistently observed on the SEC chromatographic profile (figure 4.2.3.1). Like the NTD, a higher molecular weight band was also seen on SDS-PAGE analysis of fractions from SEC of CtIP-CTD and was confirmed to be dimeric CTD by MALDI-TOF MS (Appendix supplementary figures 7.5 & 7.6). On purification of MBP-fused CTD for protein-protein interaction studies the addition of the 42 kDa MBP protein increased the separation of these peaks and allowed for purification and analysis of two distinctive populations of the CTD.

4.2.3.1 Expression and purification of MBP-CTD

The MBP-CTD construct was expressed in Rosetta™ 2 (DE3) *E.coli* cells, with soluble protein purified by sequential amylose resin and heparin chromatography stages without TEV cleavage. The sample was concentrated to 1 mL prior to final polishing by SEC using a Superdex 200 16/60 column (figure 4.2.3.1). Fractions containing either monomeric or dimeric species of CtIP-CTD were pooled separately.

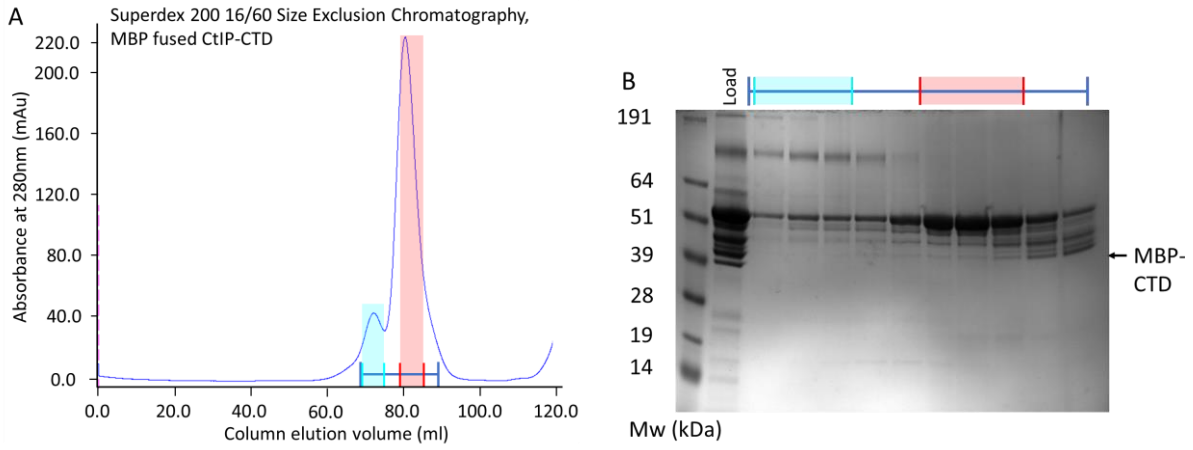


Figure 4.2.3.1: **Purification of two species of MBP fused CtIP-CTD by SEC for mass analysis.** Protein was purified using the standard procedure with the TEV cleavage stage omitted. **A)** Chromatogram of the MBP-CTD elution profile of Superdex 200 16/60 SEC. Fractions indicated were analysed by SDS-PAGE in B. **B)** SDS-PAGE analysis of protein applied (Load) to the SEC column followed by SEC elution fractions. Fractions corresponding to either monomeric (red) or dimeric (cyan) species of CtIP-CTD were pooled separately.

4.2.3.2 SEC-MALS on two CTD species

To confirm the molecular weights of the suspected monomeric and dimeric species, SEC-MALS was performed. Light scattering is a powerful technique to measure the molecular weight of a native protein without relying on protein standards. When a MALS photometer is used in combination with SEC, the molecular mass of a protein sample can be determined from analysis of its scattering characteristics coupled with accurate determination of its concentration as measured by its refractive index. The upstream SEC functions to remove contamination, aggregation, and to fractionate polydisperse materials prior to light scattering analysis. MALS has the advantage of being independent of protein shape, thus gives a more accurate estimate of protein sizes in their native oligomeric state.

The molecular masses of CtIP-CTD dimer and monomer were determined using SEC-MALS. Analysis of the monomer fraction (figure 4.2.3.1 second peak red fractions) showed a measured value of 59.2 kDa (figure 4.2.3.2a). The predicted molecular weight of MBP fused CtIP-CTD is 60.23kDa and so confirms the presence of a monomer species as the value was within the range of the instrument. On application of the CtIP-CTD dimer fraction (figure 4.2.3.1, first peak cyan fractions) to the SEC-MALS instrument two peaks were apparent with baseline separation. Molecular weight analysis of the peaks gave weights of 115.8 kDa and 59.1 kDa (figure 4.2.3.2b). The measured value for the higher molecular weight species is 1.92x that of the predicted monomeric weight and is within the range of the instrument for a dimeric species, thus showing that the CTD can form a dimer in solution.

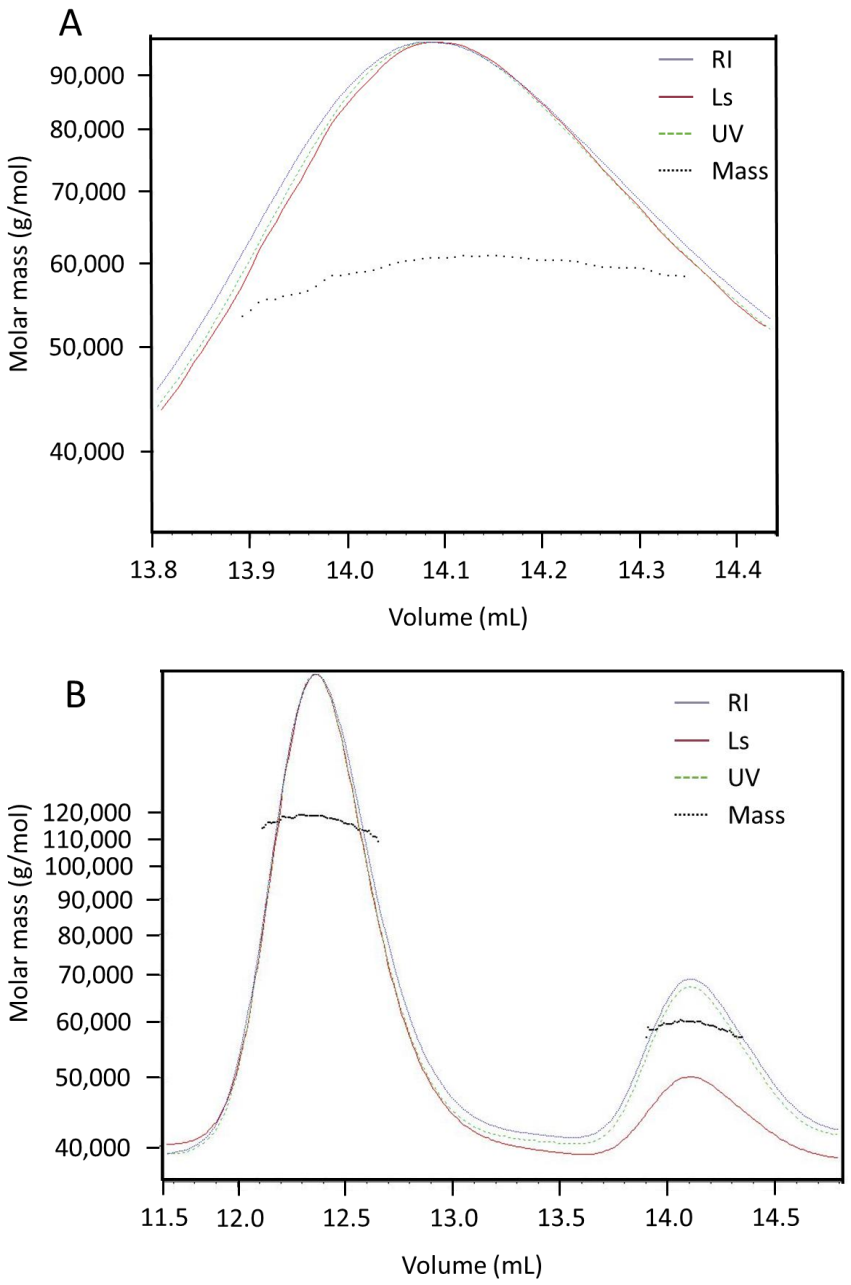


Figure 4.2.3.2: **MALS analysis of MBP fused CtIP-CTD shows monomer and dimer.** Experiment was carried out using a 10/300 Superdex 200 SEC column connected to the MALS instrumentation. RI = refractive index, Ls = light scattering, UV = UV from the SEV elution, Mass = molar mass of the peak. **A)** Analysis of MBP fused CtIP-CTD monomer fraction highlighted in red in figure 4.2.3.1 measured mass is 59.2 kDa. Polydispersity is 1.001 ± 0.126%. **B)** Analysis of MBP fused CtIP-CTD dimer, highlighted in cyan in figure 4.2.3.1 measured mass of peak 1 is 115.8 kDa with a Polydispersity of 1.000 ± 0.096%. Peak 2 is measured at 59.15 kDa with a Polydispersity of 1.000 ± 0.159%.

4.2.3.3 Zinc analysis of the dimeric CTD.

SEC chromatograms of CTD expressed using the new purification strategies showed a higher molecular weight peak, corresponding to dimeric CtIP-CTD, persisting even without zinc addition (figure 4.2.1.4-2). The initial CtIP-CTD amylose purification where zinc was added upon cell lysis (figure 4.2.1.4-2, red trace) was pooled as a monomer and dimer fraction. These two species were analysed by PAR assay with similar protein monomer: zinc ratios (figure 4.2.3.3-1). In this experiment 50 µM of dimer had a 35 µM zinc content (2.8:1 protein monomer to zinc), while 100 µM of monomer had 43 µM of zinc content (2.4:1 protein to zinc content).

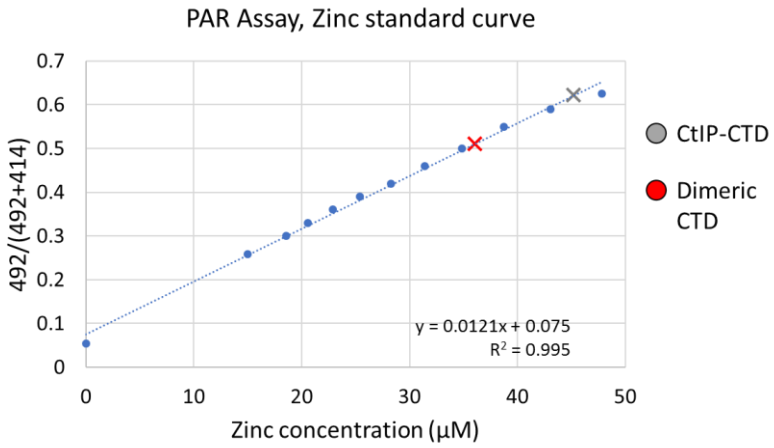


Figure 4.2.3.3-1: **Determination of zinc content for dimeric CtIP-CTD**. Blue circles represent values used in the linear fit with equation displayed. Grey cross on the line is the ratio of absorbance values at 492 and 414 nm for the CtIP-CTD sample. Red cross on the line is the ratio of absorbance values at 492 and 414 nm for the dimeric CtIP-CTD sample.

Purification optimisations to increase zinc content using zinc supplemented 2xYT media still gave a dimeric peak on SEC analysis. PAR assays were also performed on these samples with the dimeric CTD having the same 1:1 monomer protein to zinc ratio as CtIP-CTD (figure 4.2.3.3-2). In this experiment 17 µM dimeric CTD gave a zinc content of 37.0 µM (0.9:1 monomer CTD to zinc content ratio), while 33 µM monomeric CTD gave a zinc content of 38.4 µM (0.85:1 protein to zinc content ratio); each zinc value is averaged over three repeats.

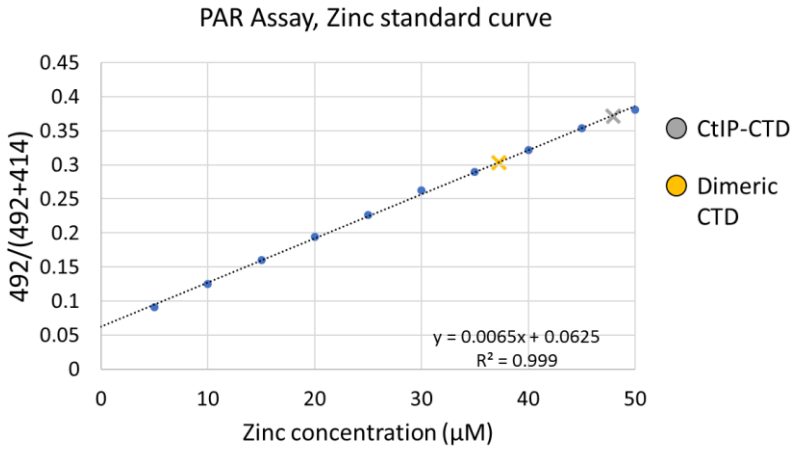


Figure 4.2.3.3-2: **Determination of zinc content for dimeric CtIP-CTD expressed using zinc supplemented media**. Blue circles represent values used in the linear fit with equation displayed. Grey cross on the line is the ratio of absorbance values at 492 and 414 nm for the CtIP-CTD sample. Yellow cross on the line is the ratio of absorbance values at 492 and 414 nm for the dimeric CtIP-CTD sample.

4.2.3.4 DNA binding of dimeric CtIP-CTD

Dimeric CTD was also tested in DNA binding assays, with FP results indicating extra zinc was still necessary for DNA binding as measured by FP. Figure 4.2.3.4-1 shows a representative trace from two independent replicates performed both requiring zinc addition.

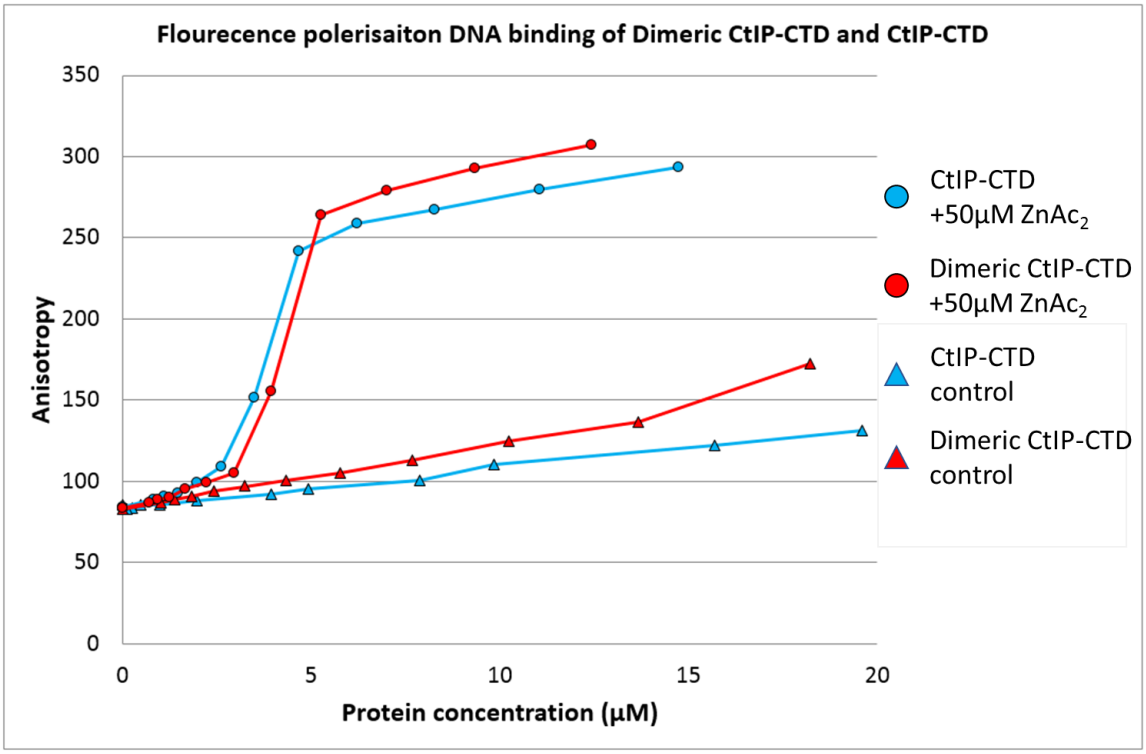


Figure 4.2.3.4-1**: Zinc dependency of FP DNA binding for Dimeric CtIP-CTD.** DNA binding with linear dsDNA fdx60. Graphical representation of dimeric CtIP-CTD binding curves with buffer containing 50 µM $ZnAc_2$ (circles) verses without zinc addition (triangles). Blue traces represent CtIP-CTD binding curves with (circles) and without (triangles) 50 µM $ZnAc_2$.

The EMSA also confirmed the interaction of the dimeric CtIP-CTD with DNA through the formation of a slow migrating species on addition of protein (figure 4.2.3.4-2). This result was repeated both with and without growth in zinc-supplemented media, with no striking difference between monomeric and dimeric protein.

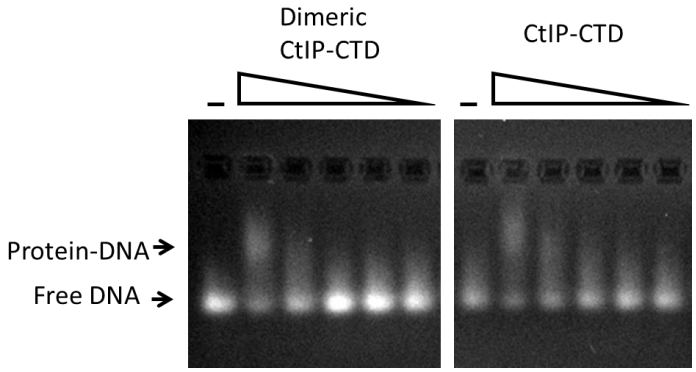


Figure 4.2.3.4-2: **Dimeric CtIP-CTD binds DNA by EMSA.** DNA binding with 1 µM dsDNA fdx60. Triangle represents decreasing amounts of CtIP-CTD or dimeric CtIP-CTD from 4.8-0.6 µM. Different protein concentrations were made by serial dilution in both experiments.

4.2.3.5 CTD_{ABC} and Threonine 847 phosphomimetic still form dimeric species

On expression of CTD_{ABC} a higher molecular weight peak is prominent on the SEC chromatogram shown by SDS-PAGE to contain CTD_{ABC}. Degradation products were also shown to co-elute. To further explore the necessity of zinc for dimer formation, semi-analytical Superdex 200 10/300 SEC runs were performed on mixed species of MBP-CTD protein. Addition of a 10-fold excess of EDTA to chelate metal ions had no effect on the amount of dimeric MBP-CTD (figure 4.2.3.5-1, P1) although it increased the levels of aggregated MBP-CTD and reduced the levels of monomeric MBP-CTD. EDTA has a k_d for zinc of approx. 10^{-16} M and is known for its ability to disrupt and aggregate zinc finger proteins such as transcription factors (Nyborg and Peersen, 2004). The observation that the dimeric species remained unchanged supports zinc being dispensable for dimer formation.

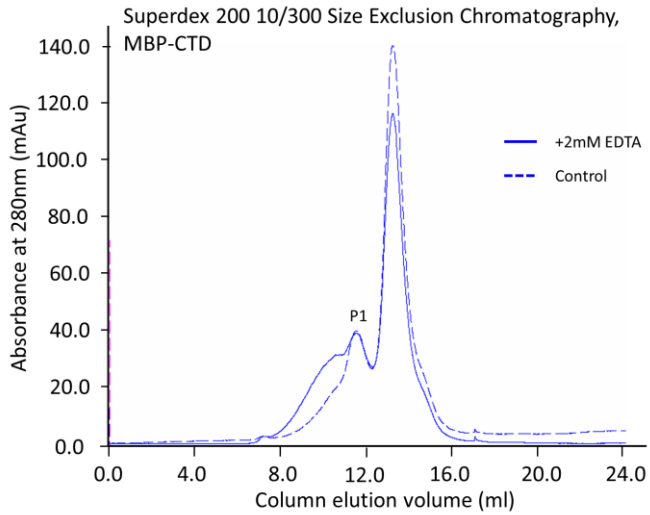


Figure 4.2.3.5-1: **Dimeric CtIP-CTD (P1) is unaffected by EDTA.** Superdex 200 10/300 SEC was performed independently on MBP-CTD both with and without the addition of 2mM EDTA to the sample prior to application to the column. All other constraints were identical.

The presence of a dimeric species in CTD_{ABC}, MBP-CTD with EDTA, and the similar protein: zinc ratios for dimeric and monomeric CtIP-CTD show that zinc is likely dispensable for dimer formation. Within the CTD the conserved threonine at position 847 has been shown to be necessary for CtIP's action in DNA repair (Sartori et al., 2007). To assess the necessity of threonine 847 of CtIP in formation of a dimeric CtIP-CTD species, residue 847 was mutated to glutamic acid (CTD_{T847E}) to mimic phosphorylation (Sartori et al., 2007). Expression and purification of CTD_{T847E} was performed from four litres of Rosetta™ 2 (DE3) cells. Upon lysis, zinc was added in accordance with the original protocol. Subsequently the amylose purification protocol was followed (figure 4.2.1.4-3), with clarified lysate applied to an amylose column to capture CTD_{T847E} before overnight dialysis incubated with TEV protease and HiTrap heparin chromatography. On SEC analysis multiple peaks were seen (figure 4.2.3.5-2a); SDS-PAGE analysis showed CTD_{T847E} to be present across all the peaks (figure 4.2.3.5-2b). The first peak contained predominantly contamination products, however, the second major peak eluted at the point of a dimeric CtIP-CTD; SDS-PAGE analysis confirmed the presence of the dimeric

species. Fractions containing either monomeric or dimeric species of CTD_{T847E} were pooled separately. The protein identity was determined by MALDI analysis to contain CtIP, further confirming the presence of a CTD_{T847E} dimer (figure 4.2.3.5-2b, blue box).

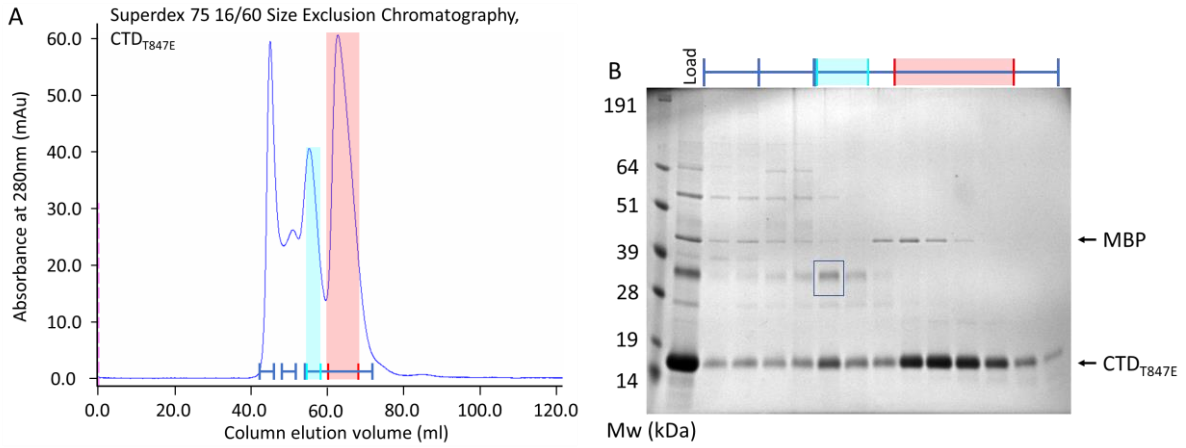


Figure 4.2.3.5-2: **Final stage purification of CTD_{T847E} by SEC. A)** Chromatogram of the CTD_{T847E} elution profile of Superdex 75 16/60 SEC with fractions analysed by SDS-PAGE in B. **B)** SDS-Page analysis of SEC fractions. Fractions indicated in red were pooled as 'CTD_{T847E}' with fractions in cyan pooled for 'dimeric CTD_{T847E}' and concentrated for further experiments.

An EMSA was performed to assess the DNA binding capability of CTD_{T847E} to determine if monomer or dimer could still bind to DNA and have any effects therein (figure 4.2.3.5-3). As noted in figure 4.2.3.5-2b, MBP was still present within the CTD_{T847E} purification and therefore protein concentrations measurements were likely overestimations. However, both CTD_{T847E} and dimeric CTD_{T847E} were able to form a complex with DNA as noted by a slower migrating species on addition of protein.

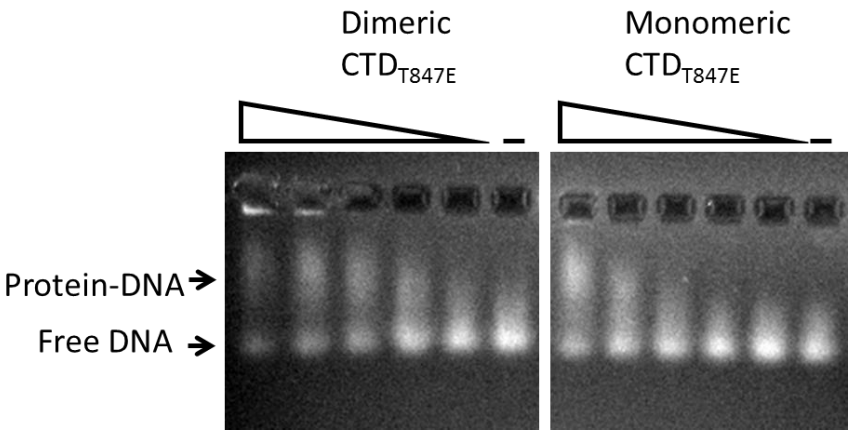


Figure 4.2.3.5-3: **CTD_{T847E} binds DNA by EMSA.** EMSA with decreasing amounts of CTD from 11-0.68 µM with 1 µM dsDNA fdx60, resulting in the formation of slow migrating species which represent protein-DNA complexes. Different protein concentrations were made by serial dilution with identical concentrations for dimeric CTD_{T847E} LHS and CTD_{T847E} RHS.

Neither changes in zinc nor a phosphomimetic T847 disrupted dimer formation or its ability to bind DNA. To further study the appearance of the dimeric species the CTD sequence was probed.

4.2.4 Presence of a coiled coil region adjacent to CTIP-CTD

Upon identification of the N-terminal coiled-coil region that was shown to form dimers by Dubin *et al*. in 2004, they speculated there was potential for a further coil-coil at the C-terminus of CtIP. Here it was reported that residues 700-775 had a 10% chance of forming a coiled coil (Dubin et al., 2004). With the apparent dimerisation of CtIP-CTD the CtIP sequence was submitted to the coiled-coil prediction software COILS (Lupas et al., 1991), MARCOILS (Delorenzi and Speed, 2002) and Paircoil2 (McDonnell et al., 2006) servers to probe the C-terminus for a further coiled-coil domain. These servers compare the sequence to an ever-expanding database of known parallel coiled-coils, therefore as new coiled-coil sequences are deposited the chance of detecting a divergent coiled-coil becomes greater. The COILS prediction highlights the N-terminal domain with the presence of two distinctive coiled-coil regions which are already known. The prediction however also highlights a second domain spanning residues 700 to 780.

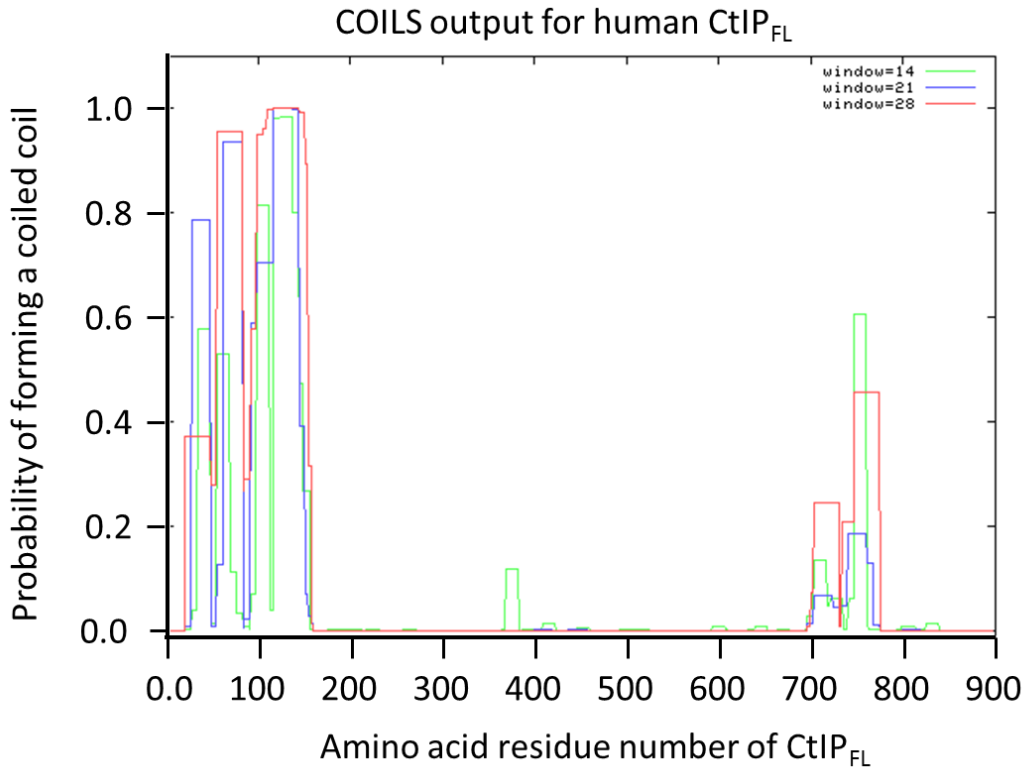


Figure 4.2.4-1: **COILS prediction for CtIP$_{FL}$ highlights a coiled-coil at both N- and C-termini.** Green line represents a search window of 14 residues, blue line represents a window of 21 residues, with red representing a window of 28 residues. A probability scale is shown on the y-axis with 1.0 indicating high probability and 0 representing low probability.

More recently developed servers MARCOIL and Paircoil2 were also performed with similar results. In particular MARCOIL at a 50% threshold predicted residues 703-774 to form a coiled-coil domain. MARCOIL also detected the NTD coiled-coil to a 90% threshold. All three of the servers predicted that the 769-domain boundary was adjacent to, and overlapping with, a potential coiled-coil region. To assess the nature of this predicted coiled-coil domain and its effect on the CTD, two further constructs were generated. Here CtIP residues 694 to 775 (dCTD) and residues 694-897 (CTDex) were cloned into pMAT11 for expression as His-MBP fusion constructs (figure 4.2.4-2). To aid purification dCTD was cloned with a C-terminal Strep tag.

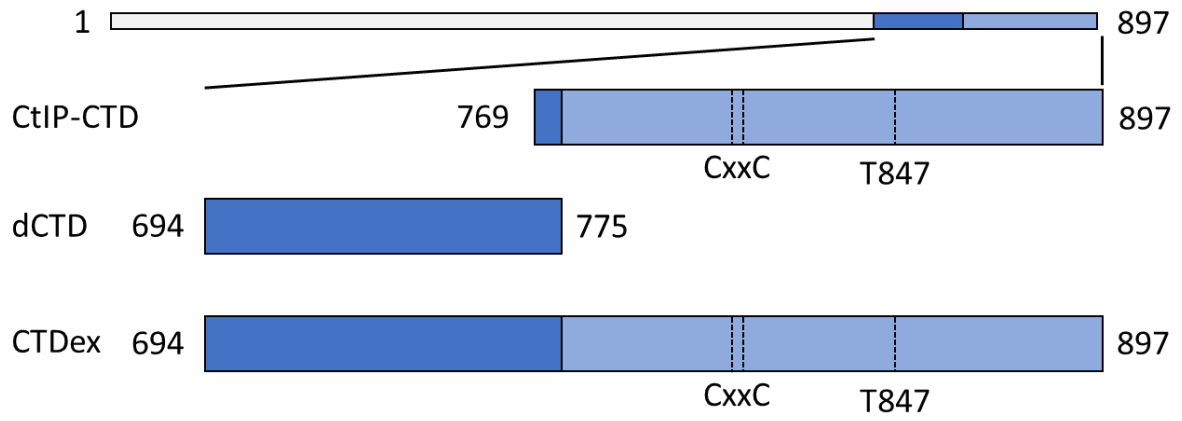


Figure 4.2.4-2: **Schematic drawing of dCTD and CTDex**. CtIP residues 694-775 are shown in dark blue, residues 775-897 are shown in light blue. Zinc binding site (CXXC) and T847 CDK site are shown.

4.2.4.1 Purification of dCTD

As dCTD is upstream of the zinc binding region (769-897) it could be purified using a Ni-NTA column procedure as metal binding was not envisaged to be an issue. In addition, the C-terminal Strep-tag aided in the removal of contaminants. Two litres of Rosetta™ 2 (DE3) cells expressing dCTD were grown using the standard procedure. Following lysis by sonication and clarification, the protein was purified using Ni-NTA prior to TEV cleavage. The C-terminal strep tag was then used for isolation of the dCTD. Figure 4.2.4.1a illustrates the results of these steps as analysed by SDS-PAGE. Unfortunately, a large proportion of dCTD didn't bind to the Strep-Tactin resin and un-cleaved MBP-dCTD persisted in the elution fraction. A final Superdex 75 16/60 SEC column resolved the fully cleaved and un-cleaved MBP-dCTD species into two separate elution volumes (figure 4.2.4.1b), as visualised by SDS-PAGE (figure 4.2.4.1c). While the first peak contained a mixture of both cleaved and un-cleaved species (possibly due to the formation of mixed-species dimers), the second peak contained homogeneous cleaved dCTD. Fractions corresponding to this latter peak were therefore pooled for concentration and further experimental analysis.

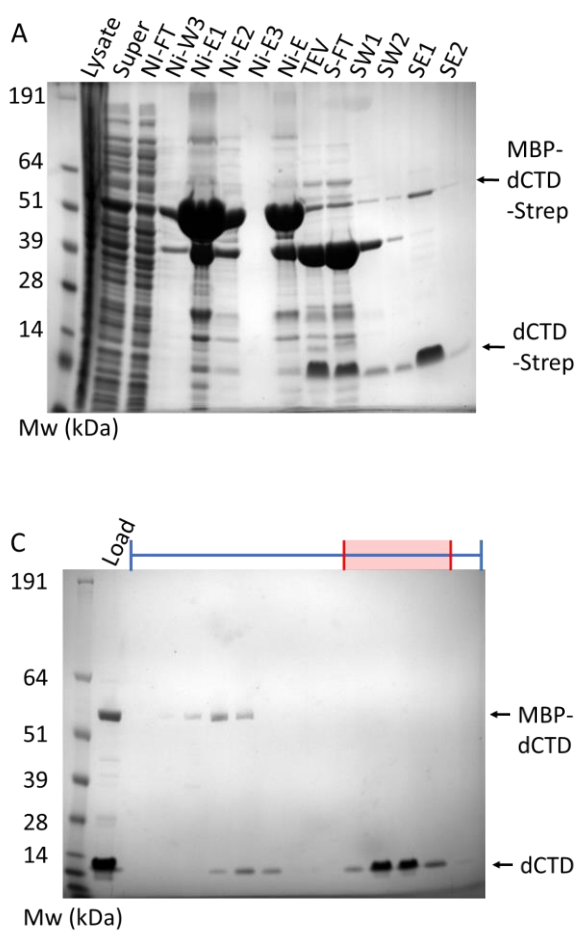


Figure 4.2.4.1: **Purification of dCTD by Ni-NTA and Strep-Tactin capture**. **A)** Initial Ni-NTA capture followed by TEV cleavage (TEV) and Strep-Tactin capture. Flow through (FT), wash (W) and elution (E) fractions are shown for Ni-NTA (N) and Strep-Tactin (S) capture stages. Ni-E1&2 were combined for TEV cleavage. SE1 was used for SEC in B. **B)** SEC trace for dCTD, fractions in blue were analysed by SDS-PAGE in C. **C)** SDS-PAGE analysis of protein loaded onto SEC along with elution fractions with MBP-dCTD and dCTD noted. Fractions in red were pooled for concentration and flash freezing.

4.2.4.2 Purification of CTDex

For the purification of CTDex an adapted amylose chromatography and heparin chromatography protocol was utilised. Two litres of Rosetta™ 2 (DE3) cells were used with standard media for the expression, with cells processed by the standard method without addition of zinc. Clarified lysate was applied to an amylose chromatography column and dialysed overnight with TEV protease. On a preliminary purification a heparin gradient was used with minimal separation of contaminants. Optimisation led to a 40% v/v buffer B step to elute and concentrate cleaved CTDex (24.4 kDa), while removing some contaminants (figure 4.2.4.2a). Protein from the first three 2 mL fractions of the 40% B elution from heparin chromatography was concentrated to 1 mL prior to a final amylose capture stage to remove further MBP and un-cleaved protein (figure 4.2.4.2b). The flow-through and wash fraction were applied to a Superdex S75 16/60 SEC column, with resulting fractions analysed by SDS-PAGE (figure 4.2.4.2c). SEC did not fully resolve the cleaved and un-cleaved CTDex species, so only central fractions from the two overlapping elution peaks were taken forward for further experiments. Optimisations to the procedure will be necessary for more homogeneous protein purification.

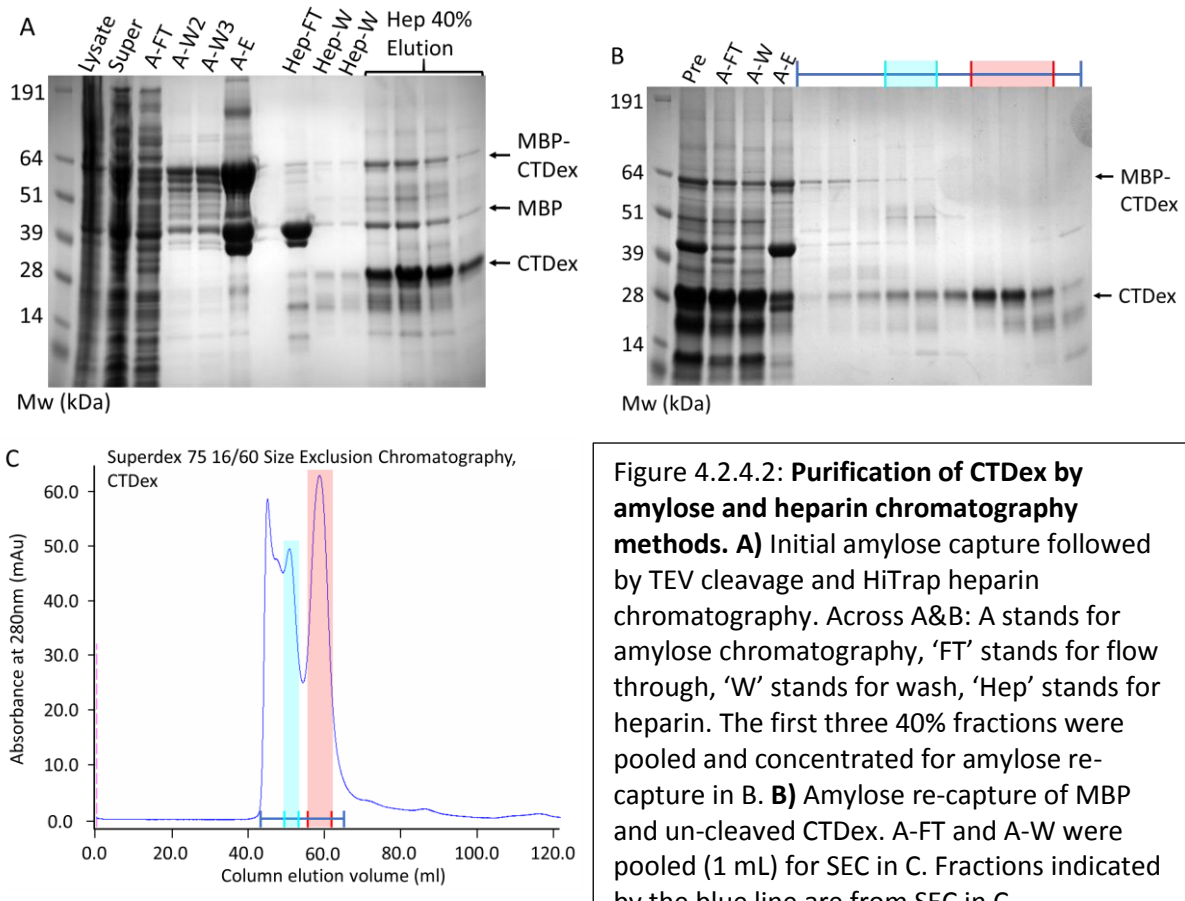


Figure 4.2.4.2: **Purification of CTDex by amylose and heparin chromatography methods. A)** Initial amylose capture followed by TEV cleavage and HiTrap heparin chromatography. Across A&B: A stands for amylose chromatography, 'FT' stands for flow through, 'W' stands for wash, 'Hep' stands for heparin. The first three 40% fractions were pooled and concentrated for amylose re-capture in B. **B)** Amylose re-capture of MBP and un-cleaved CTDex. A-FT and A-W were pooled (1 mL) for SEC in C. Fractions indicated by the blue line are from SEC in C.

C) SEC of CTDex, fractions were analysed by SDS-PAGE in B. Fractions highlighted in cyan were pooled as 'A' CTDex, with fractions in red pooled as 'B' for further analysis.

4.2.4.3 SEC analysis of dCTD and CTDex

SEC-MALS was performed at Diamond Light Source, however, there was insufficient contrast between the signal and background noise for data processing. SEC-MALS has yet to be completed using our in-house instrument due to insufficient resources. From SEC alone information about the size and shape of the protein can be determined through globular protein standards.

dCTD from its amino acid sequence has a molecular weight of 10.8 kDa, however it elutes earlier on a Superdex 75 16/60 SEC column than predicted (figure 4.2.4.3-1). From globular molecular weight standards, a protein of 17 kDa eluted at 77.3 mL on the same column. As dCTD elutes at 68.6 mL, this is a first indication that the protein is oligomeric and/or is not globular. SEC-MALS will be necessary to conclude the molecular weight of proteins within the two peaks and therefore the oligomeric nature of dCTD.

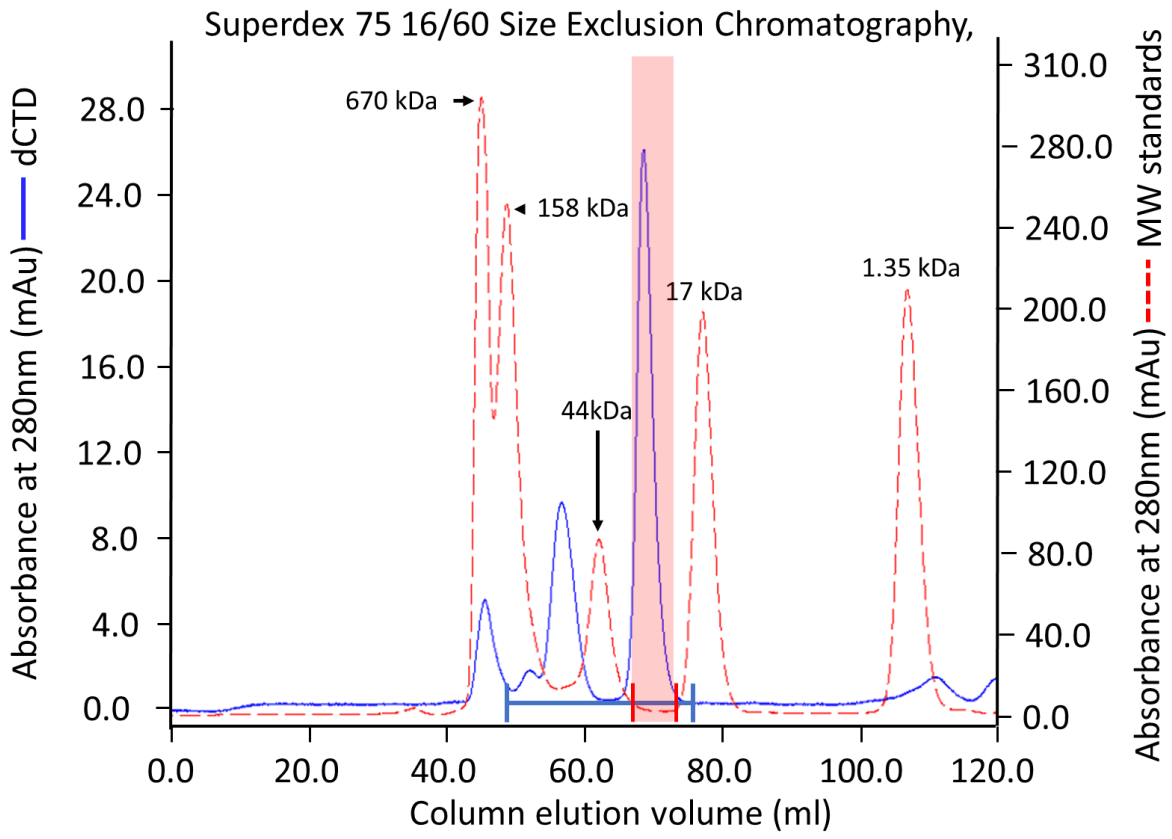


Figure 4.2.4.3-1: **Comparison between dCTD and globular molecular weight standards by SEC**. Both protein samples were applied to the same Superdex 75 16/60 SEC with the same experimental setup. Dashed red line represents Bio-Rad molecular weight standards while the solid blue line represents dCTD. Traces were overlaid using the Unicorn™ software using the elution volume. Fractions indicated by the blue line and red highlight analysed by SDS-PAGE in figure 4.2.4.1.

Analysis of CTDex using the same method also showed an earlier elution volume than expected. From the sequence the molecular weight is 24.38 kDa, thus a dimeric protein would have a molecular weight of 48.76 kDa. From SEC analysis the elution volume was 59 mL, earlier than the 44 kDa globular molecular weight standard at 62 mL (figure 4.2.4.3-2). As CTDex contains residues corresponding to the dCTD and CtIP-CTD globular molecular weights are also likely to overestimate its size, however the early elution volume does indicate the potential for oligomerisation. SEC-MALS will be necessary to confirm the molecular weight of the both proteins. SEC-SAXS was used to further probe the nature of the CTDex and dCTD (section 4.2.5) along with determination of zinc content for CTDex.

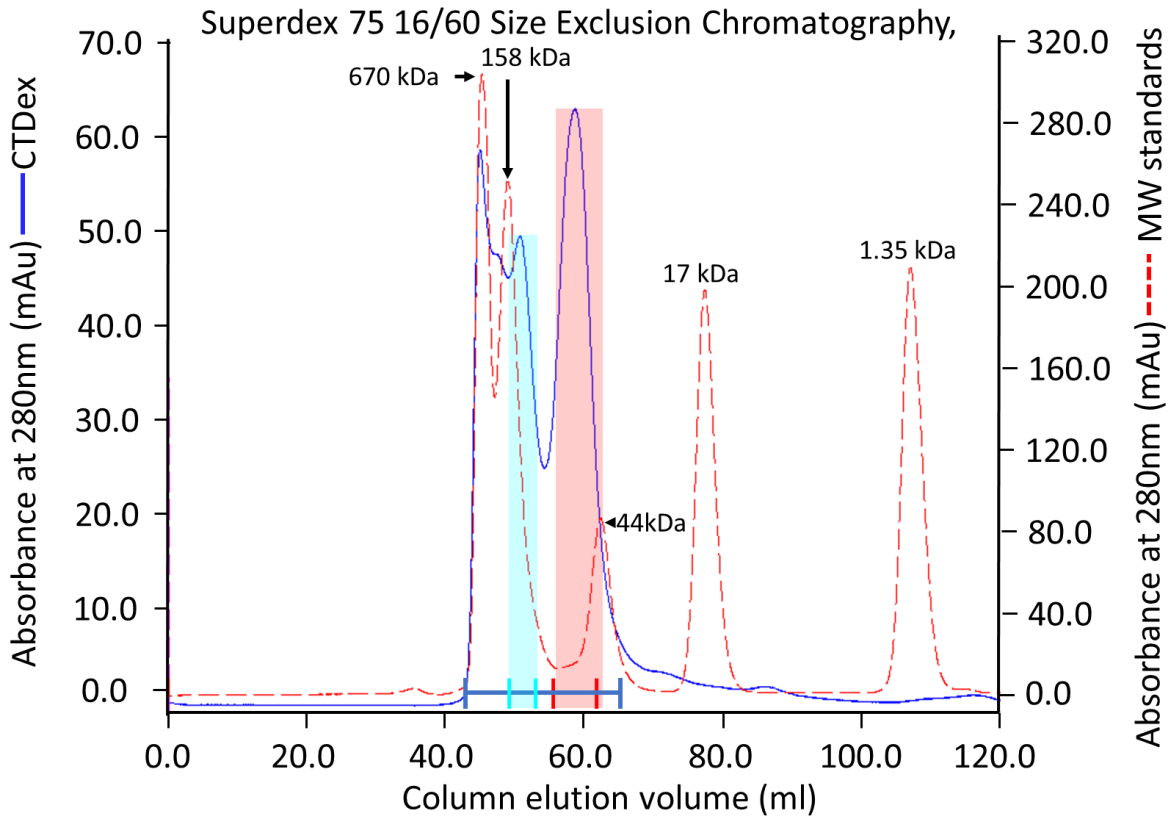
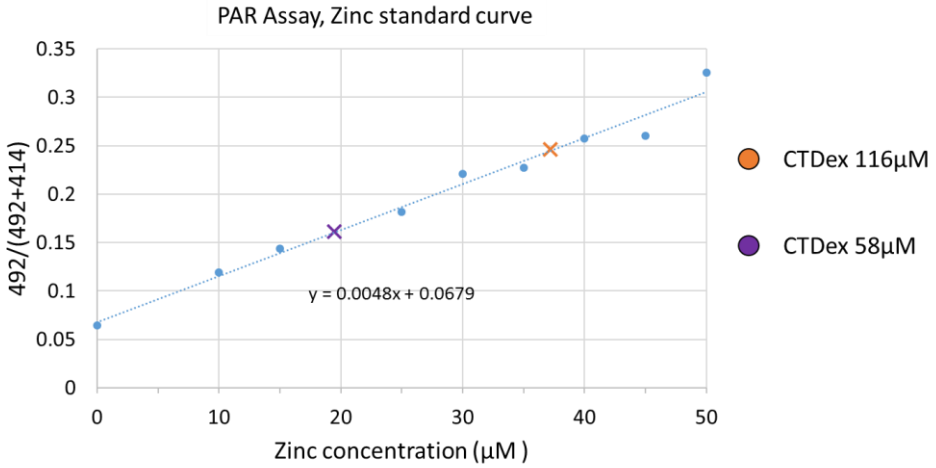


Figure 4.2.4.3-2: **Comparison between CTDex and globular molecular weight standards by SEC**. Both protein samples were applied to the same Superdex 75 16/60 SEC with the same experimental setup. Dashed red line represents Bio-Rad molecular weight standards while the solid blue line represents CTDex. Traces were overlaid using the Unicorn™ software using the elution volume. Blue and red lines show fractions analysed by SDS-PAGE in figure 4.2.4.2.

4.2.4.4 Zinc incorporation by CTDex

CTDex contains the 8-potential zinc-binding residues, therefore it should have the ability to bind zinc. Digested CTDex protein was analysed with the PAR reagent as described previously. CTDex protein assayed was purified from media without addition of further zinc using the new amylose purification method 4.2.1.4-3. Two concentrations of CTDex (CTDex fraction B figure 4.2.4.2) were digested (58 and 116 µM with respect to monomer molecular weight 24379 Da) with zinc detected through wavelength shift in peak absorbance on mixing with the PAR reagent. Analysis of triplicate assays showed CTDex to have a protein to zinc ratio of 3.0 (± 0.2):1 protein (monomer) to zinc (figure 4.2.4.4). Although this is lower than expected the zinc content is similar to dimeric CtIP-CTD from initial purifications - where zinc was only added upon cell lysis. The assay confirmed that CTDex can bind zinc without addition of zinc. Further analysis of protein purified from zinc supplemented media, and determination of oligomerisation will be necessary to confirm the ratio of CTDex to zinc bound from further purifications.



Protein	Protein concentration used/ µM	Average zinc concentration/ µM	Protein to zinc ratio
CTDex Peak B	116	35.4 ± 1.5	3.3 ± 0.33 :1
CTDex Peak B	58	19.5 ± 1.0	3.0 ± 0.2 :1

Figure 4.2.4.4: **Determination of zinc content for CTDex.** Blue circles represent values used in the linear fit with equation displayed. Orange cross on the line is the ratio of absorbance at 492 to 414 nm for the CTDex sample at 116 µM. Purple cross on the line is the ratio of absorbance at 492 to 414 nm for the CTDex sample at 58 µM. Values for 5 and 15 µM were omitted as they did not start at the baseline. Bottom panel shows the zinc concentrations and protein to zinc ratio for the two samples.

4.2.5 Low resolution structural analysis of the CtIP C-terminal domain by SEC-SAXS.

Crystallography of CtIP-CTD has yet to yield any promising avenues for structure determination. Size Exclusion Chromatography Small Angle X-Ray Scattering (SEC-SAXS) is a powerful X-ray scattering technique that can be used to obtain low resolution structural information about a biological sample in solution (Feigin and Svergun, 1989; Rambo and Tainer, 2013). Therefore, it can provide information about the shape and size of a protein without the need for crystal formation.

SEC-SAXS was used to characterise dCTD and CTDex. Normalised Kratky plots of the dCTD and CTDex gave preliminary information about their shape. The data for CTDex approximated a bell-shaped curve, which is indicative of a globular protein. Meanwhile, the data for dCTD lacked a defined inflection point, thereby indicating that this sample possesses a peptide-like or coiled-coil structure (figure 4.2.5-1).

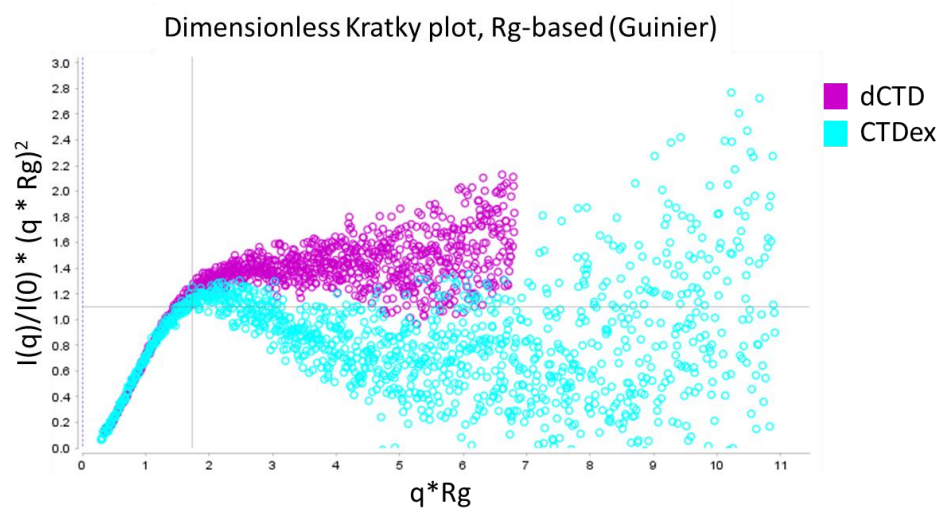


Figure 4.2.5-1: **SEC-SAXS Dimensionless Kratky plot analysis of dCTD (purple), and CTDex (cyan).**

The shape of the Dimensionless Kratky plot for dCTD was very similar to that seen for the NTD in chapter 3. Following further processing (figure 4.2.5-2) models were generated to fit the dCTD SAXS data. The dCTD model looks long and thin resembling a coiled coil (figure 4.2.5-3a) more similar to the NTD rather than to a globular protein. Comparison between the dCTD SAXS model and that of NTD_{31-145}Strep shows that dCTD can easily be positioned into the longer NTD coiled-coil with little excess envelope (figure 4.2.5-4). This supports existing evidence that dCTD exists as a coiled-coil. Comparison between the two proteins from an end-on perspective shows there is little extra room for a second molecule of dCTD within the NTD envelope further suggesting the protein measured is already dimeric. A model to fit the CTDex data shows a similar structure to dCTD at one side with a more globular head domain adjacent (figure 4.2.5-3b). These likely represent the coiled-coil and dimeric head region of the CtIP-CTD however further analysis will be necessary.

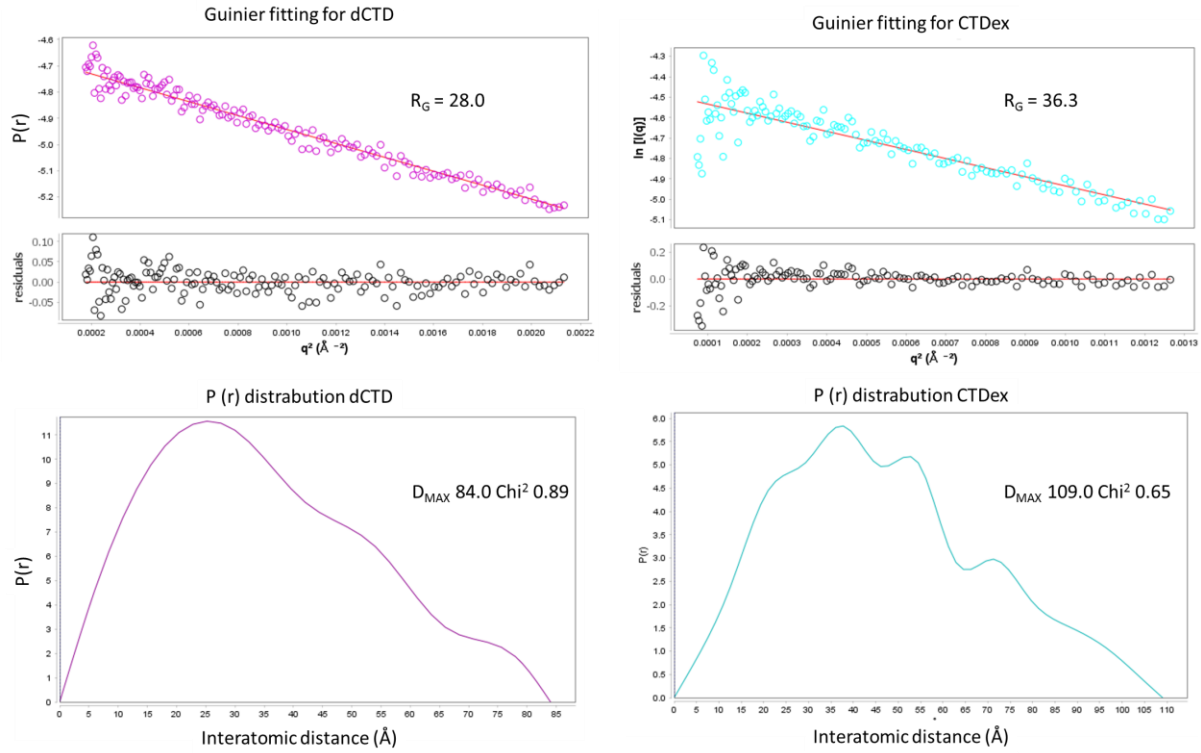


Figure 4.2.5-2: **SEC-SAXS analysis of dCTD and CTDex – Guinier fitting and P(r) distribution.** Top two panels show Guinier fitting for dCTD and CTDex with R_G displayed. Bottom two panels show the P(r) distributions for dCTD and CTDex with D_{MAX} values displayed.

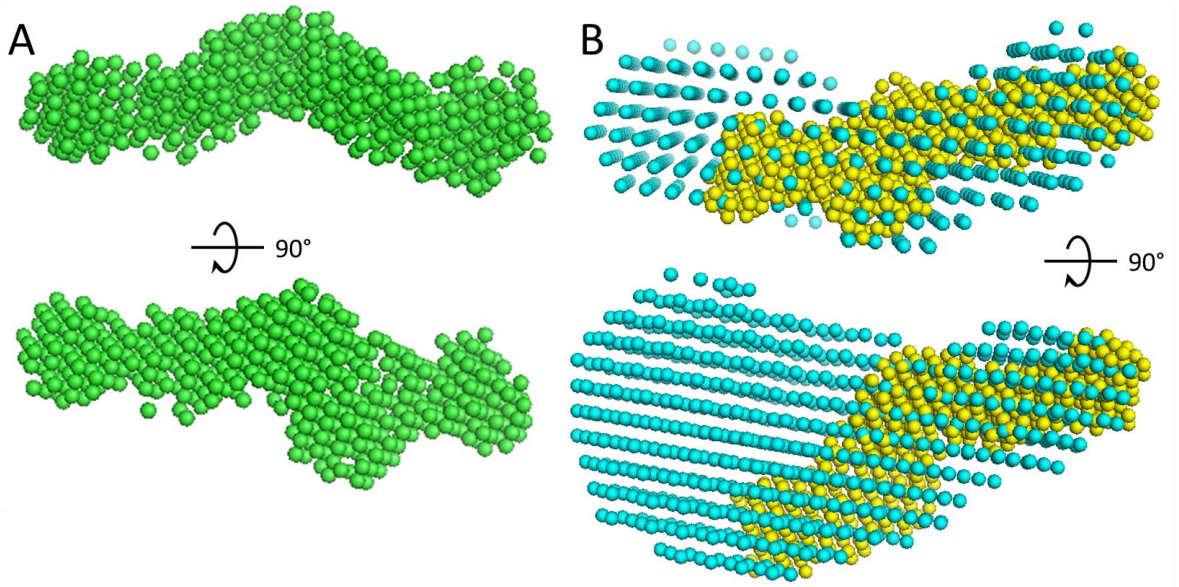


Figure 4.2.5-3: **Models to fit the SEC-SAXS data for dCTD and CTDex**. Overlay of DAMFIT models generated using DAMMIF from the SAXS data **A)** dCTD model in green shows a long rather than a globular shape. **B)** Overlay between the model for the dCTD (yellow) and CTDex (cyan). Both models were generated using DAMMIF and images displayed in PyMol.

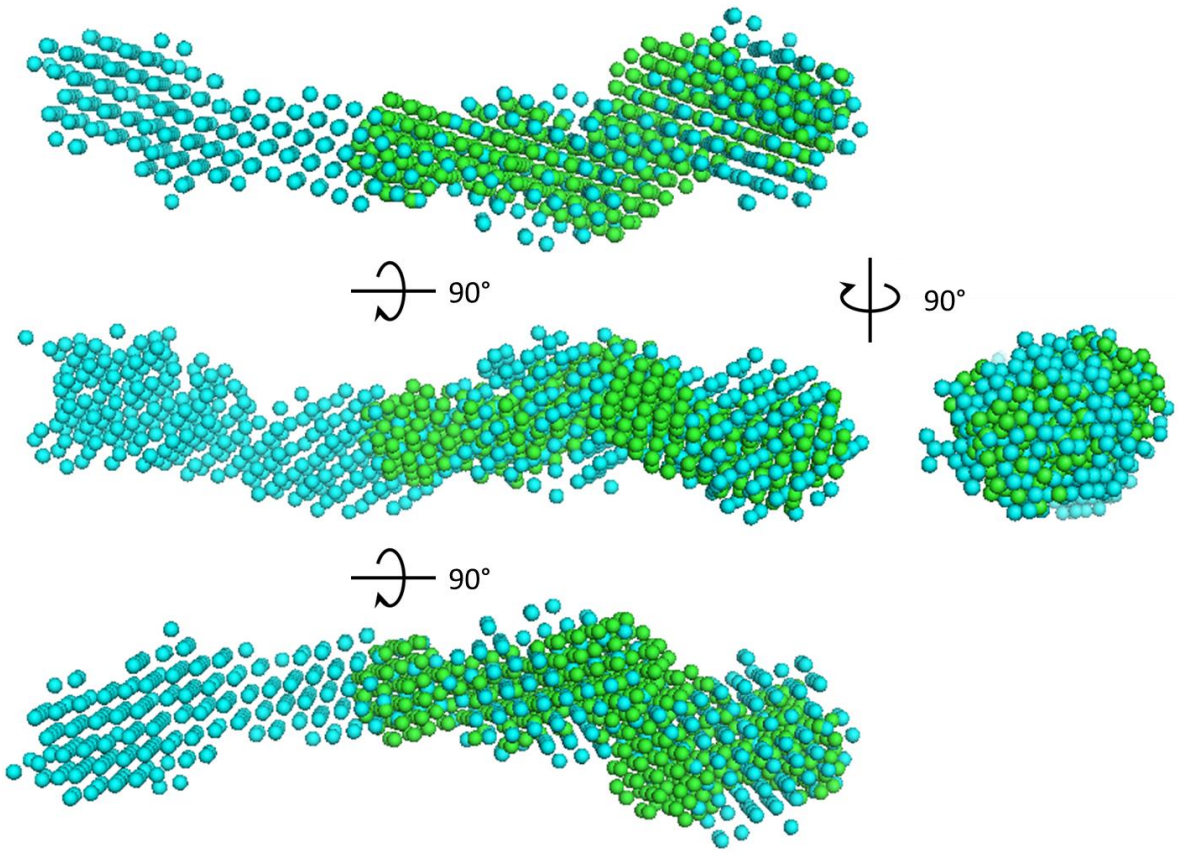


Figure 4.2.5-4: **Models to fit the SEC-SAXS data for dCTD and NTD_{31-145}Strep.** DAMFIT models generated using DAMMIF from the SAXS data for dCTD (green) and NTD_{31-145} Strep (Cyan). Models were overlaid using PyMol.

4.2.6 The CTD and its interactions with And1

Within a study by Chen *et al.* interactions were shown between And1 and CtIP by co-immunoprecipitation from HEK293T chromatin-free cell extracts (Chen et al., 2017). Deletion studies within CtIP highlighted the importance of both the N- and C-terminal domains, with the N-terminal interaction with full length And1 confirmed using purified proteins. And1 deletion studies showed residues 850-1050 formed the minimal interaction region with CtIP. However, direct interactions between And1 residues 850-1050 and CtIP or the CtIP C-terminus and And1 were not reported (Chen et al., 2017). For direct interaction between And1 and CtIP pulldown experiments were performed. Various And1 constructs (figure 4.2.6) were purified from either insect cells or bacteria with constructs provided by Dr Mairi Kilkenny (Kilkenny et al., 2017). An MBP-tagged CTD was purified (figure 4.2.6) as bait in pulldown experiments. The cleavage stage within the CTD purification was removed and the MBP-CTD fusion protein purified using the new amylose purification procedure to homogeneity. The CtIP-CTD 'monomer' fraction was used in all pull-down experiments (figure 4.2.1.4-3).

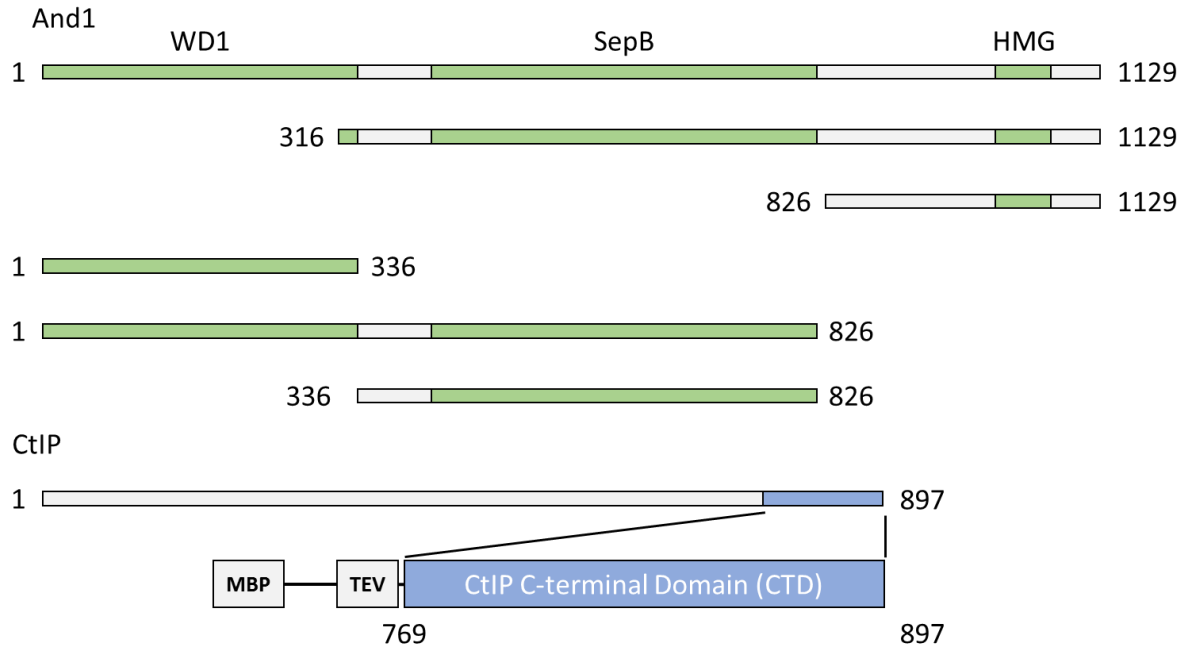


Figure 4.2.6: **Schematic drawing of And1 and MBP-CtIP-CTD used in pull-down experiments.**

4.2.6.1 Purification of full-length And1

The And1 full length ($And1_{FL}$) construct was expressed in *Sf*9 insect cells and harvested 3 days post transfection. The soluble protein was purified using sequential Ni-NTA and Heparin affinity chromatography steps. Final polishing was performed using a 16/60 Superdex 200 SEC (figure 4.2.6.1). Fractions were analysed by SDS-PAGE, with fractions containing purified And1 pooled and concentrated for pulldown experiments.

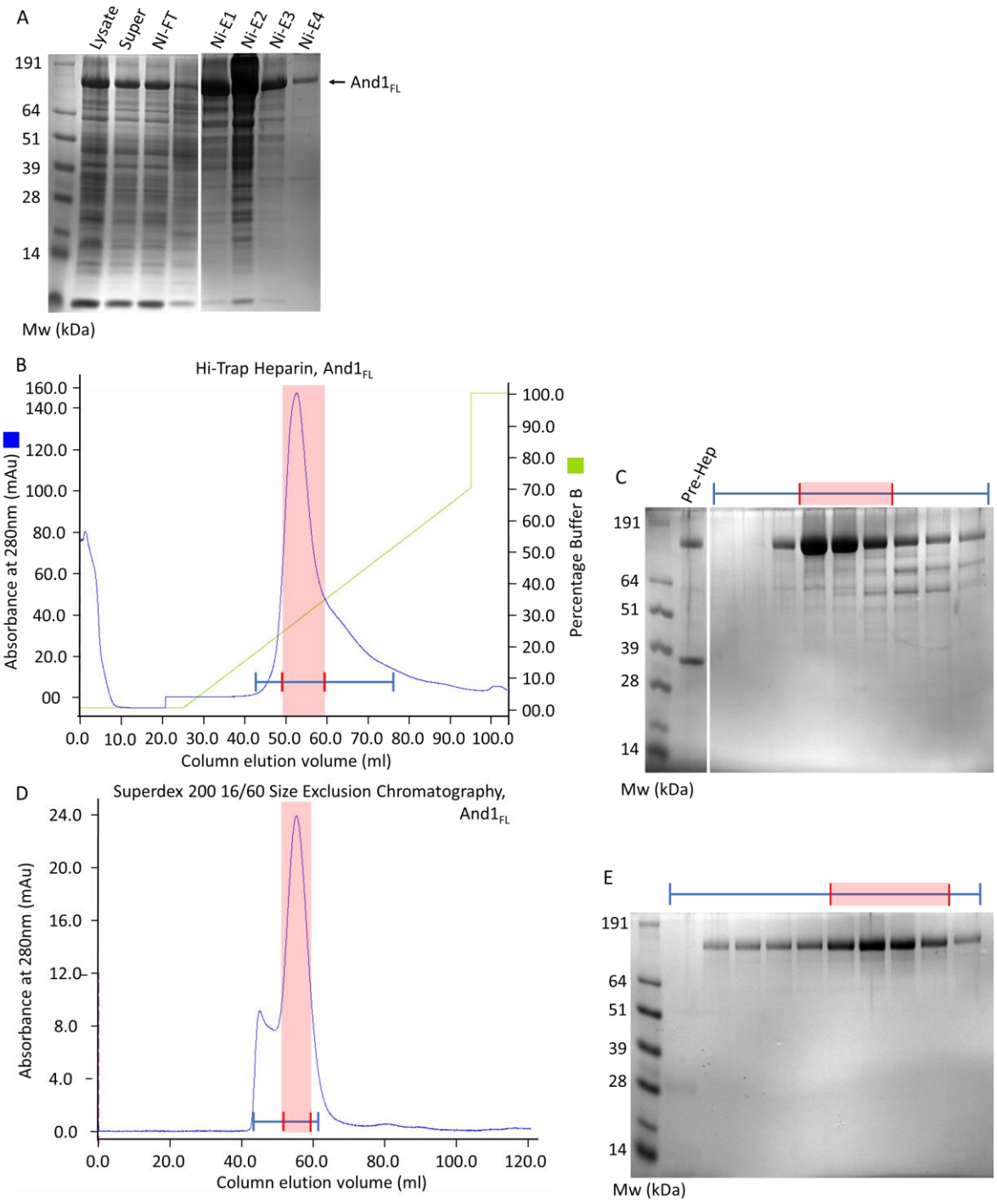


Figure 4.2.6.1: **Purification of And1 expressed in insect cells by Ni-NTA and Hi-Trap heparin. A)** Initial Ni-NTA purification with lysate, supernatant, Ni-NTA flow through and elution fractions shown. Elution fractions 1-3 were pooled for Hi-Trap heparin in B. **B)** Hi-Trap heparin chromatogram, buffer B was 1M NaCl. Fractions correspond to SDS-PAGE in C. **C)** SDS-PAGE analysis of Hi-Trap Heparin fractions with red highlight indicating fractions pooled for SEC in D. **D)** SEC chromatogram of applied And1 protein to the Superdex 200 16/60 column. Fractions correspond in SDS-PAGE in E. **E)** Analysis of SEC fractions: fractions highlighted in red were pooled and concentrated for further experimental analysis.

4.2.6.2 Interaction of And1 with CtIP CTD

To verify the And1$_{FL}$ interactions reported by Chen *et al.*, pull-down experiments using purified CtIP MBP-NTD and CtIP MBP-CTD as 'bait' vs full length And1$_{FL}$ 'prey' were performed. See chapter 2 for buffers and protocol. And1$_{FL}$ was found within CtIP-CTD elution fractions but not with CtIP-NTD.

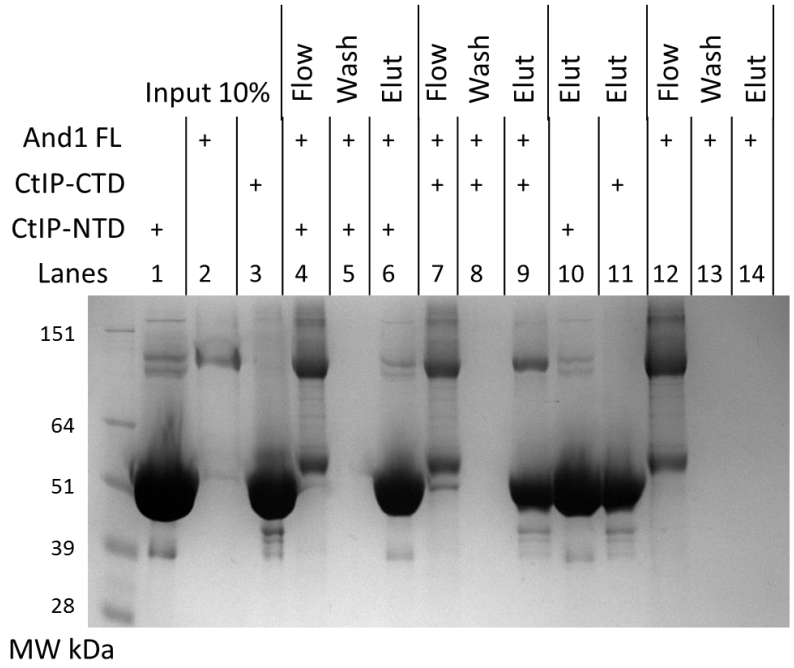


Figure 4.2.6.2-1: **CtIP-CTD, but not CtIP-NTD can retain And1 by pull-down assay**. And1 was tested for interaction with MBP-tagged CtIP-CTD or NTD bound to amylose resin. Results are shown by SDS-PAGE. And1$_{FL}$ was flowed over the resin (flow) before washing (wash) and eluting with 20 mM maltose (elution). Presence of each protein incubated with the beads is indicated by "+".

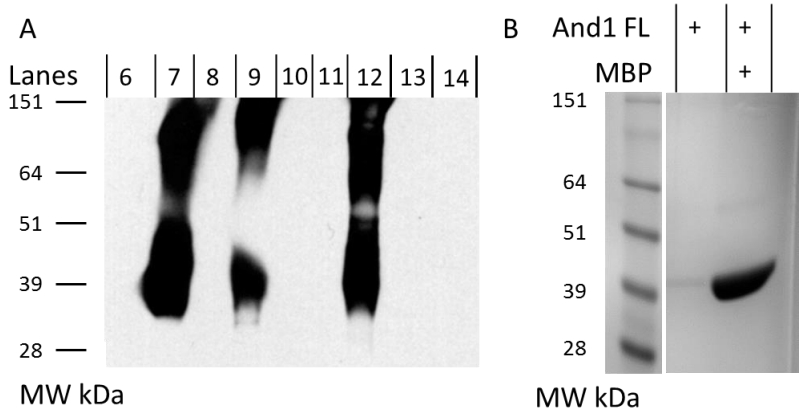


Figure 4.2.6.2-2. **Western blot analysis of CtIP-CTD and CtIP-NTD pull-down assay. A)** Western blot using anti Myc tag antibody against the And1$_{FL}$ Myc tag, lanes indicated are from figure 4.2.6.2-1. **B)** Pull-down of And1$_{FL}$ with MBP on the amylose beads.

Western Blot analysis confirmed the interaction of And1 and CtIP-CTD. No And1$_{FL}$ through virtue of its Myc tag could be detected in the presence of CtIP-NTD, thus confirming And1$_{FL}$ does not interact with CtIP-NTD. And1$_{FL}$ was unable to independently bind to the amylose resin (figure 4.2.6.2-1). Further pull-down experiments with CtIP-CTD were performed to find the And1 interaction domain.

4.2.6.3 Identification of the And1: CtIP-CTD interaction domain

To confirm which domains of And1 were necessary for the CtIP-CTD interaction, And1 constructs And1$_{316-1129}$ (no WD domain), And1$_{826-1129}$ (no WD or SepB domains) and And1$_{1-316}$ (the WD domain)

were tested. And1 proteins for these pull-downs were purified and provided by Dr Mairi Kilkenny (Kilkenny et al., 2017). The same buffers and protocols were used as for the original $And1_{FL}$ pulldown experiment. Figure 4.2.6.3-1 confirms the interaction of CtIP-CTD with $And1_{FL}$. CtIP-CTD however failed to retain $And1_{1-316}$ (figure 4.2.6.3-1 lane 20) and $And1_{826-1129}$ (figure 4.2.6.3-1 lane 12) on the beads showing that the interaction was not through the WD or C-terminal domain of And1 respectively.

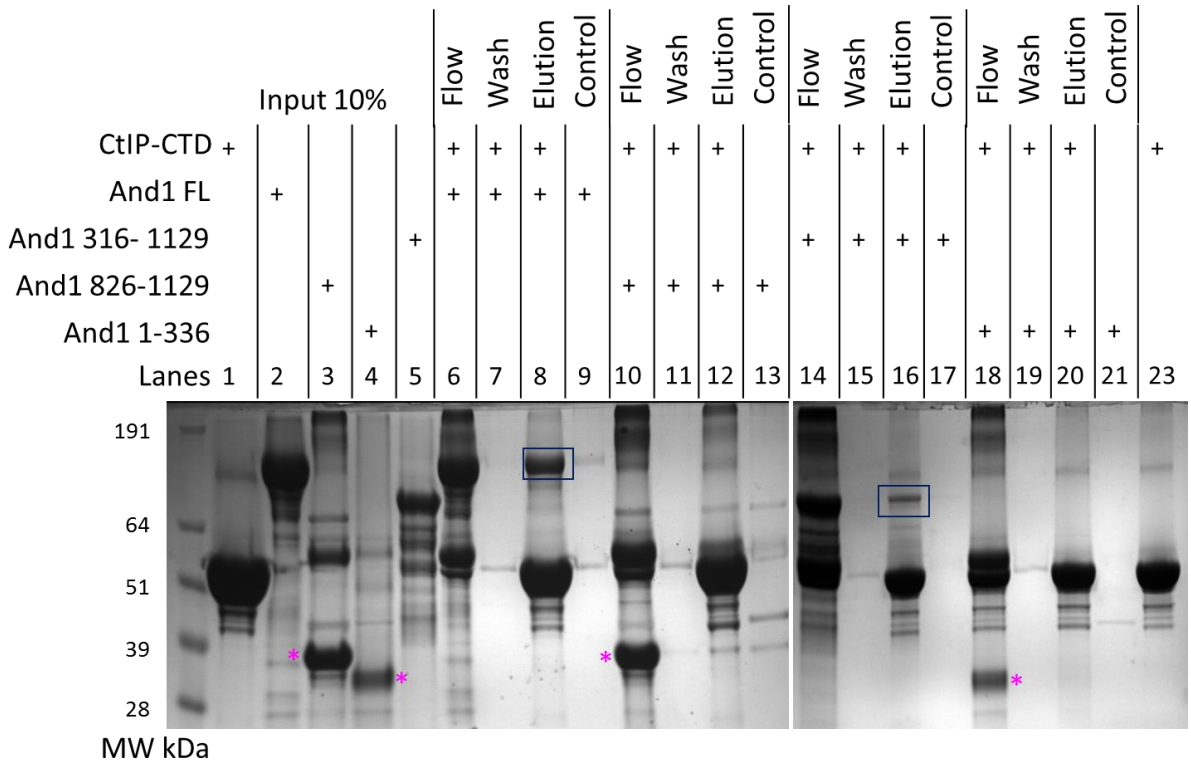


Figure 4.2.6.3-1**: And1 WD1 domain is dispensable for interaction with CtIP-CTD**. CtIP-CTD was bound to amylose beads And1 constructs $And1_{FL}$, $And1_{316-1129}$, $And1_{826-1129}$ and $And1_{1-316}$ were flowed over the resin (flow) in the pull-down experiments before washing (wash) and eluting with 20 mM maltose (elution). Presence of each protein incubated with the beads is indicated by a "+". Blue boxes show interactions, pink stars highlight non-interacting proteins.

$And1_{1-316}$ was further shown to be dispensable through $And1_{316-1129}$ being retained on the beads in the presence of CtIP-CTD (figure 4.2.6.3-1), highlighting the SepB domain from residues 316-826 as the potential interaction site. $And1_{1-826}$ containing the WD and SepB domains was purified from Rosetta™ 2 (DE3) *E. coli* for testing against CtIP-CTD. The soluble protein was purified using sequential Ni-NTA and Hi Trap Q chromatography as detailed in methods. Figure 4.2.6.3-2 shows the combined elution from the Hi Trap Q column as the input for $And1_{1-826}$. Amylose beads containing CtIP-CTD were able to retain a band of correct size for $And1_{1-826}$ within the amylose elution while $And1_{1-826}$ could not bind to the amylose resin in the absence of CtIP-CTD suggesting an interaction between the two proteins. Figure 4.2.6.3-2 also further confirms the interaction of $And1_{FL}$ with CtIP-CTD.

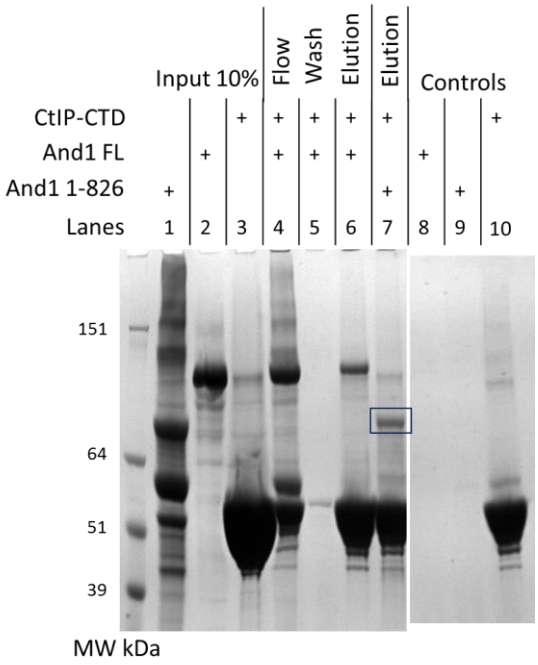


Figure 4.2.6.3-2**: CtIP-CTD interacts with And1 through its WD1-SepB domain.** CtIP-CTD was bound to amylose beads And1 constructs $And1_{FL}$, and $And1_{1-826}$ were the flowed over the resin (flow) in the pull-down experiments before washing (wash) and eluting with 20 mM maltose (elution). Presence of each protein incubated with the beads is indicated by a "+".

While MBP-tagged CtIP-CTD retained $And1_{316-1129}$ and $And1_{1-826}$ on the amylose beads, $And1_{1-316}$ and $And1_{826-1129}$ could not associate with the CtIP-CTD under the same conditions. This focuses the interaction of CtIP with And1's SepB domain, rather than the WD domain or HMG box at the C-terminus.

So far pull-down experiments have narrowed down the interaction between And1 and CtIP to the SepB domain of And1 and CTD of CtIP. Therefore, a construct spanning And1 residues 316-826 ($And1_{316-826}$) was expressed and purified from bacterial cells. The soluble protein was purified using sequential Ni-NTA and Hi Trap Q chromatography, with a SEC 16/60 Superdex 200 gel filtration final polishing stage (figure 4.2.6.3-3). Fractions were confirmed by SDS-PAGE to contain $And1_{316-826}$, however it was also noted that $And1_{316-826}$ had the same molecular mass by SDS-PAGE as MBP fused CtIP-CTD (figure 4.2.3.1). Consequently, a direct observation of an interaction using the same pull-down protocol was not possible (figure 4.2.6.3-4).

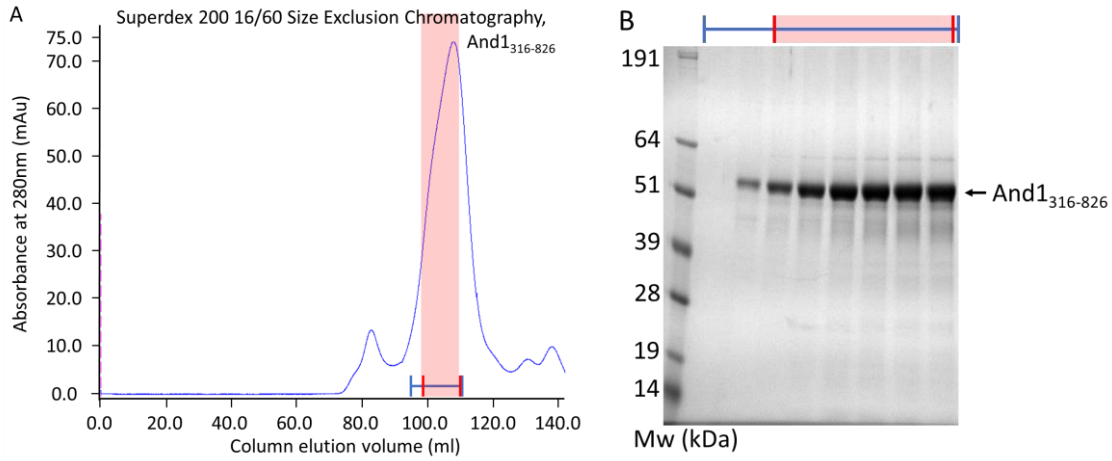


Figure 4.2.6.3-3: **Final stage purification of $And1_{316-826}$ by SEC**. **A)** Superdex 16/60 SEC chromatogram with fractions corresponding in B. **B)** fractions analysed by SDS-PAGE from SEC in A. Fractions highlighted in red were pooled for experimental analysis.

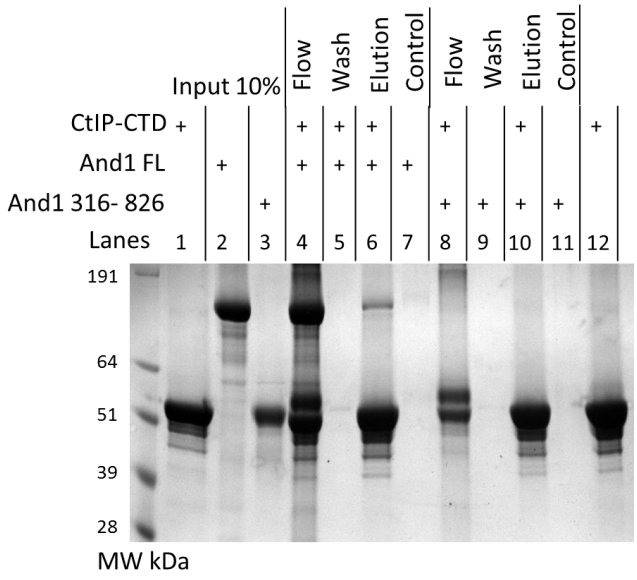


Figure 4.2.6.3-4**: CtIP-CTD pulldown experiment with $And1_{316-826}$.** CtIP-CTD was bound to amylose beads And1 construct $And1_{FL}$ and $And1_{316-826}$ were the flowed over the resin (flow) in the pull-down experiments before washing (wash) and eluting with 20 mM maltose (elution). Presence of each protein incubated with the beads is indicated by a "+".

To overcome this obstacle the protocol was adapted to include a TEV digestion stage post elution of protein from the amylose beads (see chapter 2) to cleave the MBP tag from the CtIP-CTD and reveal any proteins bound of a similar size. This allowed the MBP tag to isolate CtIP-CTD on the beads for interaction studies while later removing the overlapping band. Using this method, the MBP-CTD band is degraded into a 42 kDa MBP and 15 kDa CtIP-CTD band along with the 28 kDa band for the TEV protease present, (figure 4.2.6.3-5 lane 14). And1$_{FL}$ interaction was confirmed for this pull-down experiment with the presence of And1$_{FL}$ not affected by TEV digestion (figure 4.2.6.3-5 lane 5&6). Post TEV digestion of the And1$_{316-826}$ pull down elution a band of 51 kDa corresponding to the molecular weight of And1$_{316-826}$ was now apparent (figure 4.2.6.3-5 lane 11). Western blot analysis using an antibody against His-tagged proteins confirmed that the band at 51 kDa (figure 4.2.6.3-5 boxed) was not an un-digested MBP tagged CtIP-CTD (figure 4.2.6.3-5 western lane 11). MALDI-TOF mass spectrometry was also used to confirm that the band (figure 4.2.6.3-5 boxed) contained And1.

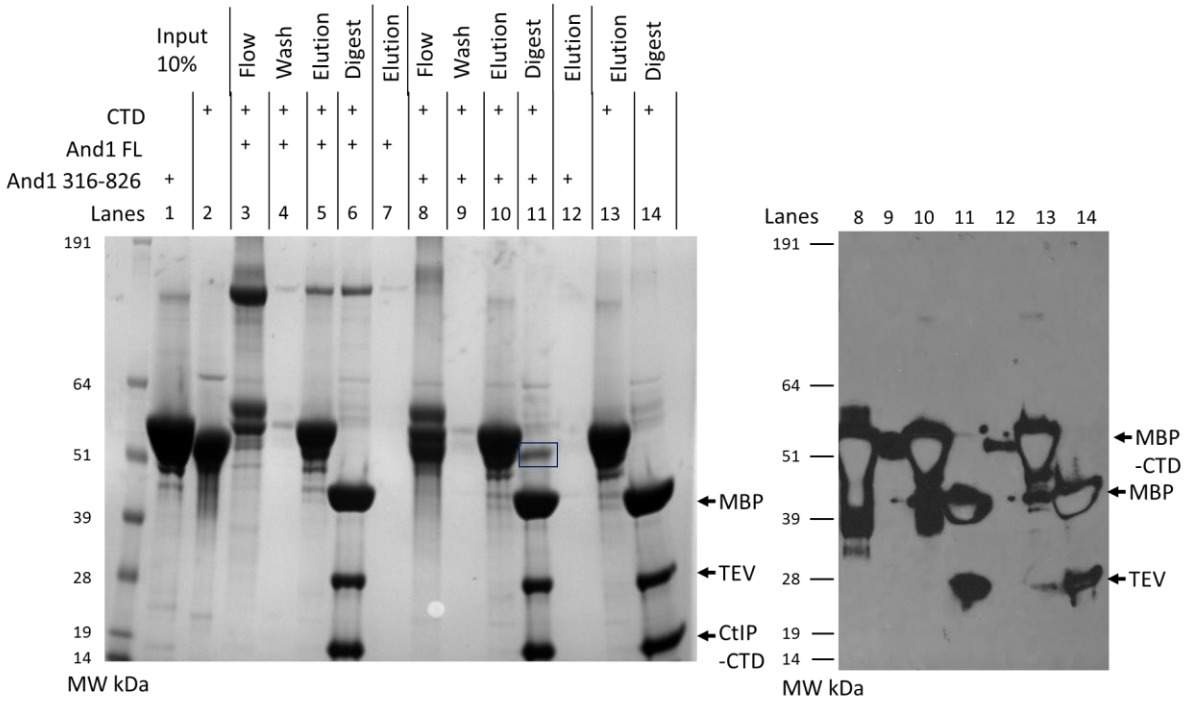


Figure 4.2.6.3-5**: CtIP-CTD interacts with And1$_{316-826}$ SepB domain**. CtIP-CTD was bound to amylose beads, And1 constructs And1$_{FL}$, and And1$_{316-826}$ were flowed over the resin (flow) in the pull-down experiments before washing (wash) and eluting with 20 mM maltose (elution). Digest is post 1-hour incubation with TEV protease with location of MBP, TEV and CtIP-CTD digestion products indicated. Presence of each protein incubated with the beads is indicated by a "+". LHS of image shows a western blot analysis of the indicated lanes using an anti-his antibody to detect his tagged proteins such as non-cleaved MBP-CTD, His-MBP tag, and His-tagged TEV.

4.3 Conclusions and future work

Sequence homology within the C-terminal region of human CtIP is highly conserved across a wide range of eukaryotic species. This, along with the drastic phenotypes apparent in Seckel syndrome, suggests a highly important role in CtIP function. The results presented here have provided insights into this region. Secondary structure analysis of CtIP predicts a coiled-coil motif near the C-terminus, juxtaposed to the functionally conserved homology domain. Chromatography techniques have been exploited to investigate the oligomeric state of the broader C-terminal region, including analytical SEC (using several different conditions), SEC-MALS and SEC-SAXS. The results of these techniques all support the existence of a dimeric species. SEC-SAXS generated an elongated envelope for this region, supporting the presence of a coiled-coil upstream of the CtIP Sae2 homology domain at the C-terminus. Future work should include further SEC-MALS and AUC experiments with dCTD and CTDex to confirm the oligomeric state.

Further to this, these domains are proposed as regions of interest for crystallographic study. The dCTD is of particular interest as its predicted coiled-coil structure is likely to be the most crystallographically tractable component of this region. Protein purification for the dCTD currently produces homogeneous protein for one of the two species. An additional step to remove any un-cleaved MBP-tagged protein would allow both species to be analysed by SEC-MALS to determine the target for crystallisation. For CTDex, a change of approach may be necessary; addition of a C-terminal strep-tag might aid purification of homogeneous protein.

The C-terminal region contains 8 putative zinc-binding Cys/His residues. The ability of these residues to bind zinc was assessed in this study with no single residue abolishing zinc binding. Further mutational analyses will be required to determine the contribution of each residue to zinc-protein complex formation. Alternative methods for zinc incorporation during protein purification have also brought into question the residues involved in zinc binding. Zinc was found to be important for the stabilisation of monomeric species of the CTD. Notably, protein aggregation was observed when either metal chelators were added, or zinc supplementation was omitted during purification. Since putative zinc-binding residues are broadly distributed across the domain, it is plausible that zinc coordination stabilises weak structural elements within the whole region. However, dimeric species of the CTD do not show the same sensitivity to zinc loss, indicating that dimerisation helps to stabilise the domain.

Biophysical studies have revealed that CtIP-CTD has an affinity for DNA in the low micromolar range, with a minimal interaction length of 20 to 25 bp. While this is not considered a high-affinity interaction, it may be modulated by dimerisation and/or post-translational modification of the CTD. It is also

plausible that the CTD undergoes induced folding upon binding to other protein factors, to further increase its affinity for DNA. Further DNA-binding experiments will be required to elucidate the effect of dimerisation, zinc and phosphorylation on this DNA-CtIP interaction.

To probe CtIP's involvement with the replication fork the interaction between the replication fork-associated protein And1 and CtIP was explored. This interaction had been previously reported through co-immunoprecipitation analysis (Chen et al., 2017). However, here alternative domains have been identified using purified And1 and purified CtIP proteins with an interaction between $And1_{316-826}$ and CtIP-CTD shown within this thesis. The different interaction sites determined between the two studies supports existence of alternate forms of CtIP, with interactions and conformations within the protein affecting the C-terminal domain. This would explain the differences seen between CtIP-CTD expressed alone verses removal of the CtIP-CTD or CtIP-NTD by Chen *et al*.

As the crystal structure of And1 residues 424-824 has been determined (Kilkenny et al., 2017), the binding of the CtIP-CTD to $And1_{316-826}$ provides a potential avenue for determination of a CtIP-And1 protein complex crystal structure. Future work within this project is necessary to identify the And1-CtIP interaction motif though the overarching aim will be to crystallise and determine the structure of the complex. Methods such as Hydrogen-Deuterium Exchange (HDX) may be used to probe the interaction surfaces between And1 and CtIP, helping to guide crystallisation efforts. Interactions between other repair proteins (such as MRN) and the protein constructs described in this chapter should also be explored.

Chapter 5, CtIP Full-Length

5.1 Introduction

CtIP is a protein fundamental to life with ubiquitous cellular expression and a multitude of different roles from transcription to DNA repair. Apart from several short motifs that are characterised as protein-protein interaction sites and a large array of reported post translational modifications, the only known structural element of the protein is a coiled-coil domain at the N-terminus of CtIP (Dubin et al., 2004; Fusco et al., 1998). Later structural studies have demonstrated the presence of a tetramerisation domain preceding and joined to the coiled-coil region. The impact of abolishing tetramerisation on HR repair is known through cellular assays, showing a strong defect on DNA-end resection (Davies et al., 2015).

Functional studies of CtIP have mainly been limited to cell-based investigations due to the difficulty of recombinant protein purification. Where research groups have expressed CtIP (Anand et al., 2016; Liao et al., 2016; Makharashvili et al., 2014; Sartori et al., 2007; Wang et al., 2014) this has been mainly for radiation-based gel assays with no biophysical or structural analysis of full length CtIP published to date. Structure prediction from the protein sequence can provide an avenue for biophysical characterisation of sub domains, however for CtIP this is limited to its N- and C-terminus. Therefore, a large portion (~600 amino acids or 66%) of the protein remains unexplored. With potential for N-terminal modifications to affect the C-terminal of the protein, research into CtIP as an intact recombinant protein is therefore paramount. In order to characterise structurally human CtIP, I set out to express and purify the full-length protein. This chapter reports the successful production of recombinant CtIP and its biochemical and biophysical characterisation.

5.2 Results

5.2.1 *E. coli* expressed $CtIP_{FL}$

E. coli is the workhorse of recombinant protein expression and is the expression system of choice due to its low cost, ease of manipulation and high protein yields. Since C-terminal and N-terminal $CtIP_{FL}$ constructs had been shown to require the MBP tag for expression and solubility, $CtIP_{FL}$ was cloned in frame with an MBP tag as a first expression test using Image cDNA clone 4830693. PCR of the $CtIP_{FL}$ ORF required annealing temperature optimisation to remove contaminants (figure 5.2.1.1b). The $CtIP_{FL}$ open reading frame was cloned into pMAT11 using restriction sites *Sal*I and *Nco*I, generating a His_6MBP tagged CtIP construct.

5.2.1.1 Bacterial protein expression and purification of Human $CtIP_{FL}$

Expression trials of the $CtIP_{FL}$ construct were carried out using the Rosetta™2 (DE3) strain. Figure 5.2.1.1a shows the result of the Ni-NTA purification from 2 litres of cells. No protein band attributable to the 139 kDa His_6MBP-$CtIP_{FL}$ was visible by SDS-PAGE or by Anti-His Western blot (figure 5.2.1.1c). As the bacterial expression was not successful, an alternative expression system was required.

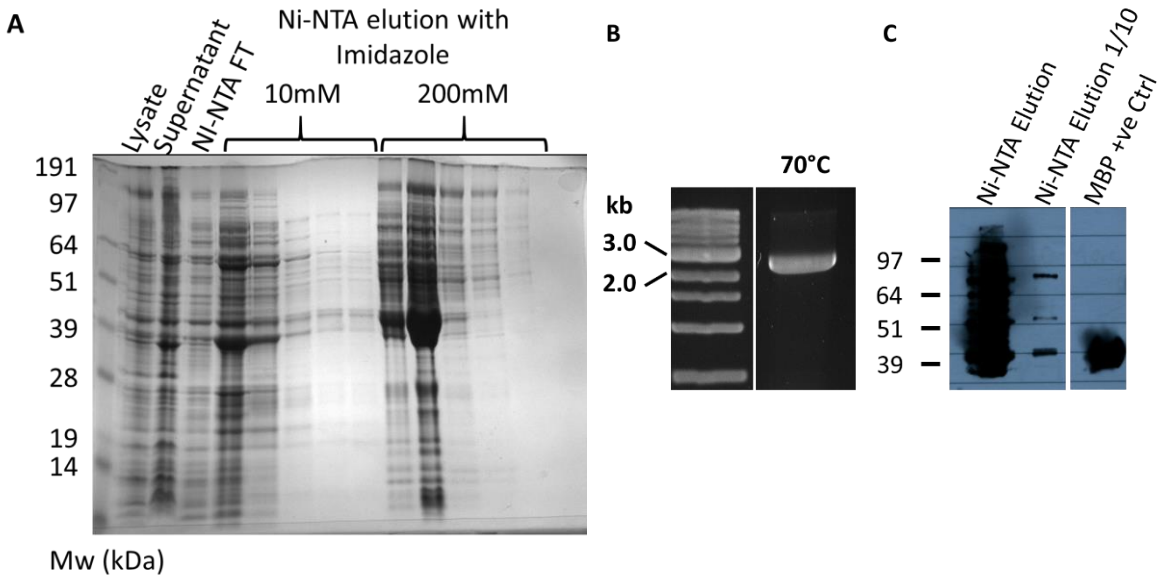


Figure 5.2.1.1: **Cloning and expression trial of $CtIP_{FL}$ expressed in E. coli.** Image clone 4830693 was amplified, cloned and expressed in Rosetta™ 2 (DE3) bacterial cells. **A)** Result of Ni-NTA affinity purification. 'FT' stands for flow-through. Combined elution fractions 1-3 were analysed by Western blot in C. **B)** PCR product from optimised $CtIP_{FL}$ PCR reaction, 70°C was found to be the optimum annealing temperature. **C)** Western blot analysis of Ni-NTA elution fractions using an anti-His tag antibody that can detect His-MBP tagged $CtIP_{FL}$.

5.2.2 Insect-cell expression and purification of $CtIP_{FL}$

Baculovirus expression in insect cells is a well-established system that is suitable for expressing large amounts of eukaryotic protein, and has been previously deployed with success by other research groups to produce recombinant $CtIP_{FL}$ (Makharashvili et al., 2014). The Flag-tagged $CtIP_{FL}$ construct donated by Tanya Paull was used for baculovirus generation.

5.2.2.1 $CtIP_{FL}$ expression test

Following p2 virus generation, a small-scale expression test was performed infecting 50 mL of *Sf*9 cells at $2x10^6$ cells per mL. At four time intervals (+1 day, +2days, +3days, +4 days) 1 mL of cells were removed for expression testing. Cells were pelleted and suspended in lysis buffer: 20 mM Tris pH 8, 300 mM NaCl, 0.5 mM TCEP, 10% v/v glycerol and SigmaFAST™ protease inhibitor tablets EDTA free. Total lysate and supernatant of each time point were analysed by Western blot using an Anti-Flag antibody to detect the Flag-tagged $CtIP_{FL}$ protein. Expression of $CtIP_{FL}$ was detected from +2 days onwards, together with evidence of degradation products (figure 5.2.2.1). From this experiment it was therefore determined that cells should be harvested at +2 days to ensure a maximum $CtIP_{FL}$ yield and minimise the formation of degradation products.

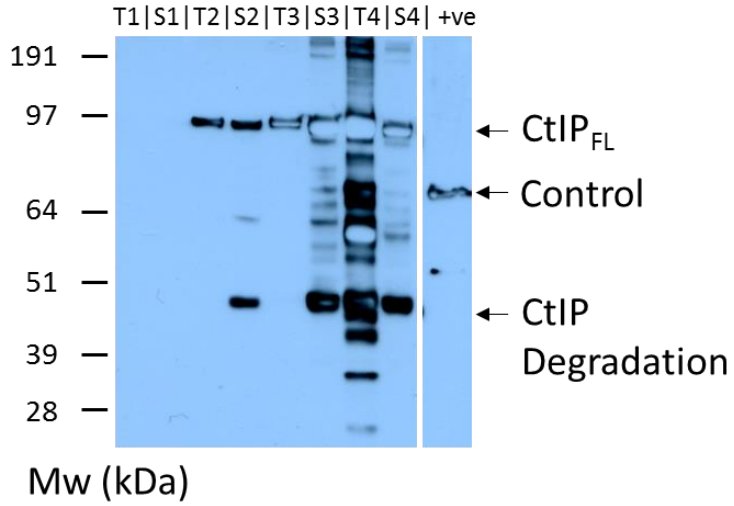


Figure 5.2.2.1: **Western blot of $CtIP_{FL}$ *Sf9* expression test using Anti-flag antibody.** Samples corresponding to total protein (T) and Supernatant (S) from clarified lysate are shown on the Western blot. 'Control' is the flag tagged Mre11.

5.2.2.2 $CtIP_{FL}$ protein purification

500 mL of *Sf*9 cells infected with Flag-tagged $CtIP_{FL}$ baculovirus were grown and harvested at +2 days. Filtered cell lysate supernatant was first passed over a 1 mL anti-Flag tag resin column, washed, and the retained Flag-tagged $CtIP_{FL}$ eluted with Flag-peptide buffer (figure 5.2.2.2a). Fractions were analysed by SDS-PAGE and pooled accordingly; the $CtIP_{FL}$ was then concentrated to 0.5 mL and applied

to a Superdex 200 10/300 SEC column pre-equilibrated into buffer containing 20 mM Tris pH 8, 300 mM NaCl, 0.5mM TCEP and 10% v/v glycerol. Eluted fractions containing $CtIP_{FL}$ as shown by SDS-PAGE were pooled and concentrated to 400 µL. Protein was snap frozen with liquid nitrogen in 40 µL aliquots. The bands from SDS-PAGE analysis (figure 5.2.2.2b) were identified by MALDI-TOF mass spectrometry (PNAC facility, department of Biochemistry) as $CtIP_{FL,}$ Hsp70, and a degradation product of CtIP.

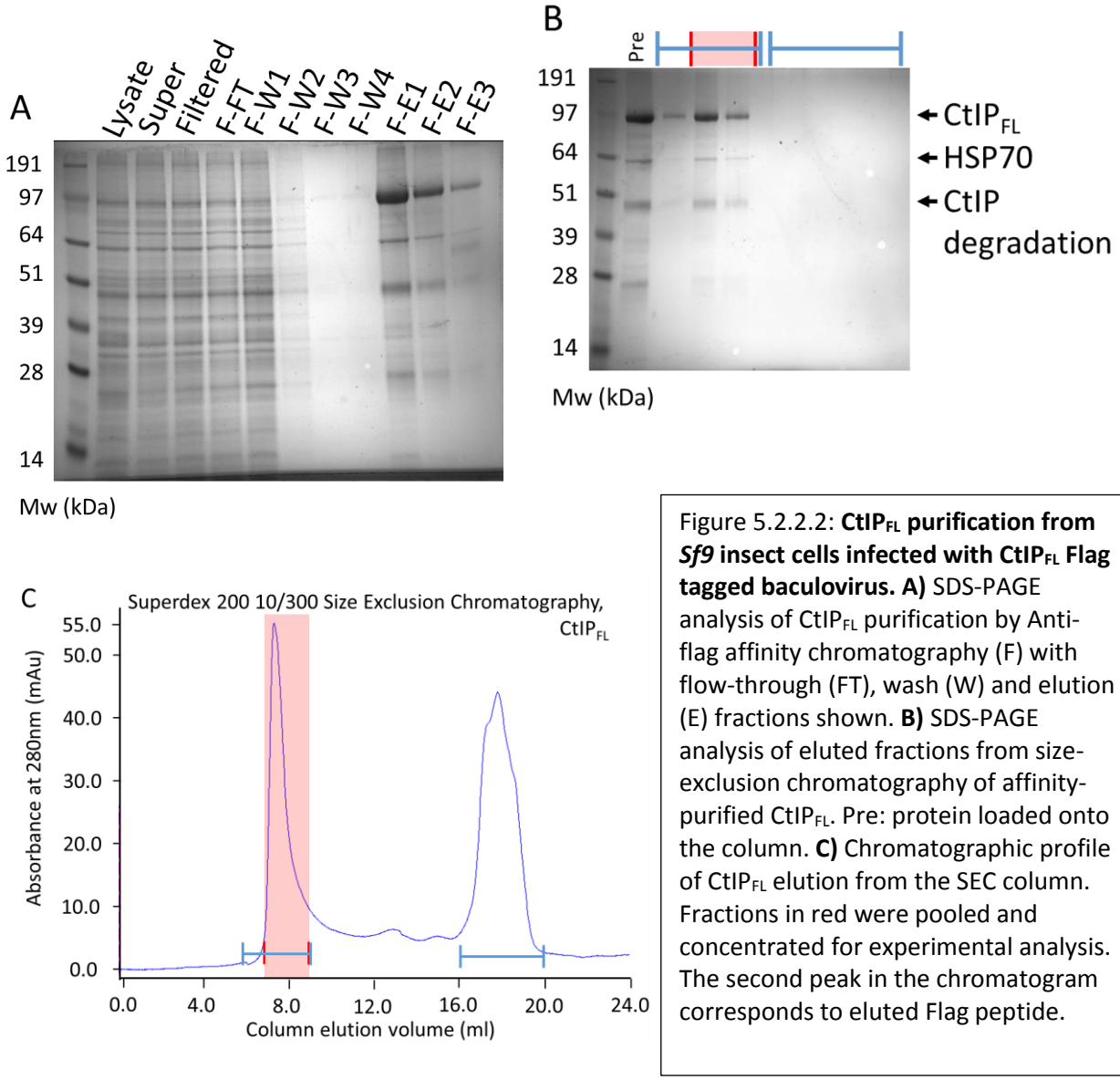


Figure 5.2.2.2: **$CtIP_{FL}$ purification from *Sf9* insect cells infected with $CtIP_{FL}$ Flag tagged baculovirus. A)** SDS-PAGE analysis of $CtIP_{FL}$ purification by Anti-flag affinity chromatography (F) with flow-through (FT), wash (W) and elution (E) fractions shown. **B)** SDS-PAGE analysis of eluted fractions from size-exclusion chromatography of affinity-purified $CtIP_{FL}$. Pre: protein loaded onto the column. **C)** Chromatographic profile of $CtIP_{FL}$ elution from the SEC column. Fractions in red were pooled and concentrated for experimental analysis. The second peak in the chromatogram corresponds to eluted Flag peptide.

The concentration of purified and concentrated $CtIP_{FL}$ was measured as 0.7 mg/mL or 6.75 µM (monomer), using a theoretical extinction coefficient of 41715 M^{-1}. Given the presence of a contaminant species as HSP70 and of $CtIP_{FL}$ degradation products, the concentration value is likely to represents an over-estimate of the true concentration of $CtIP_{FL}$ in the sample.

5.2.2.3 CtIP$_{FL}$ DNA binding

Human CtIP and its fission-yeast orthologue S. *pombe* Ctp1 have been shown to possess DNA binding ability, residing within their conserved C-terminal domains (Andres et al., 2015; Davies et al., 2015). Therefore, the DNA-binding ability of recombinant human CtIP$_{FL}$ was tested by EMSA alongside that of CtIP-CTD (figure 5.2.2.3).

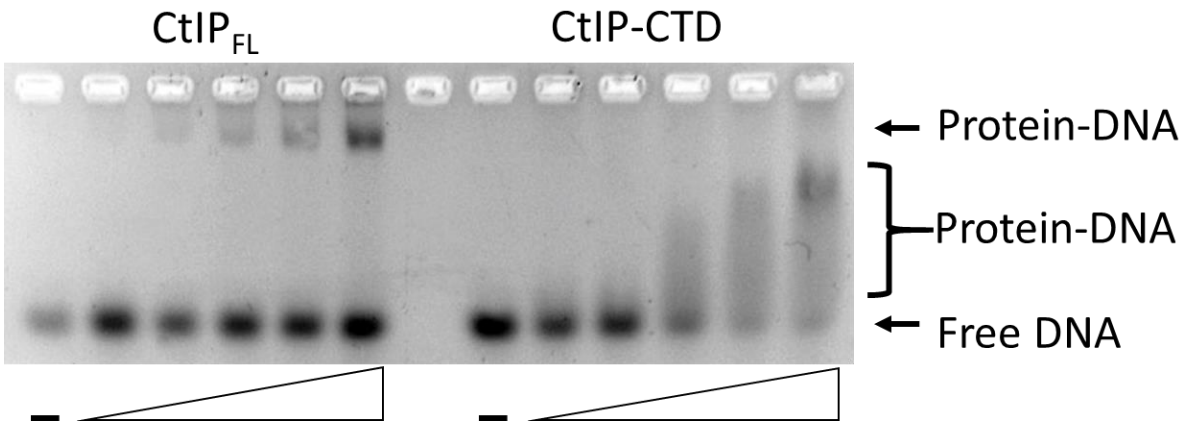


Figure 5.2.2.3: **DNA binding of insect expressed CtIP$_{FL}$ by EMSA.** Left-hand side: EMSA of 1 µM 40 bp dsDNA in the presence of CtIP$_{FL}$ at concentrations of 6 µM, 3 µM, 1.5 µM, 0.75 µM, and 0.375 µM. Right-hand side: EMSA of 1 µM 40 bp dsDNA in the presence of CtIP-CTD at concentrations of 12 µM, 6 µM, 3 µM, 1.5 µM and 0.75 µM. The position in the gel of the free dsDNA and the slower migrating protein-DNA complexes are indicated.

Clear differences in DNA binding behaviour between CtIP$_{FL}$ and CtIP-CTD were apparent. The CtIP$_{FL}$-DNA complex migrated as a tight band, unlike the CtIP-CTD-DNA species that showed a more diffuse pattern in the gel, a possible indication that this complex is unstable during electrophoresis. In contrast, CtIP-CTD was able to shift almost all the DNA probe, although the inability of CtIP$_{FL}$ to achieve a similarly complete bandshift might be due to an overestimation of its protein concentration. The observed differences in the electrophoretic mobility behaviour of the protein-DNA complexes formed by CtIP$_{FL}$ and its CTD prompted me to explore the role of CtIP oligomerisation in DNA binding.

5.2.2.4 L27E CtIP$_{FL}$ protein purification

The mutation of leucine residue 27 to glutamic acid (L27E) has been shown to abrogate tetramerisation (Davies et al., 2015). Therefore, if tetramerisation of CtIP$_{FL}$ is important for DNA binding, a L27E CtIP mutant should have reduced or no DNA binding ability. To generate the L27E mutation on CtIP$_{FL}$ the Quikchange mutagenesis protocol was used on the CtIP$_{FL}$ Flag-tagged pACEBac1 construct. After DNA sequence confirmation of the L27E mutation, a L27E CtIP$_{FL}$ bacmid and baculovirus were generated. L27E CtIP$_{FL}$ was expressed and purified in the same way as the wild-type protein (figure 5.2.2.4).

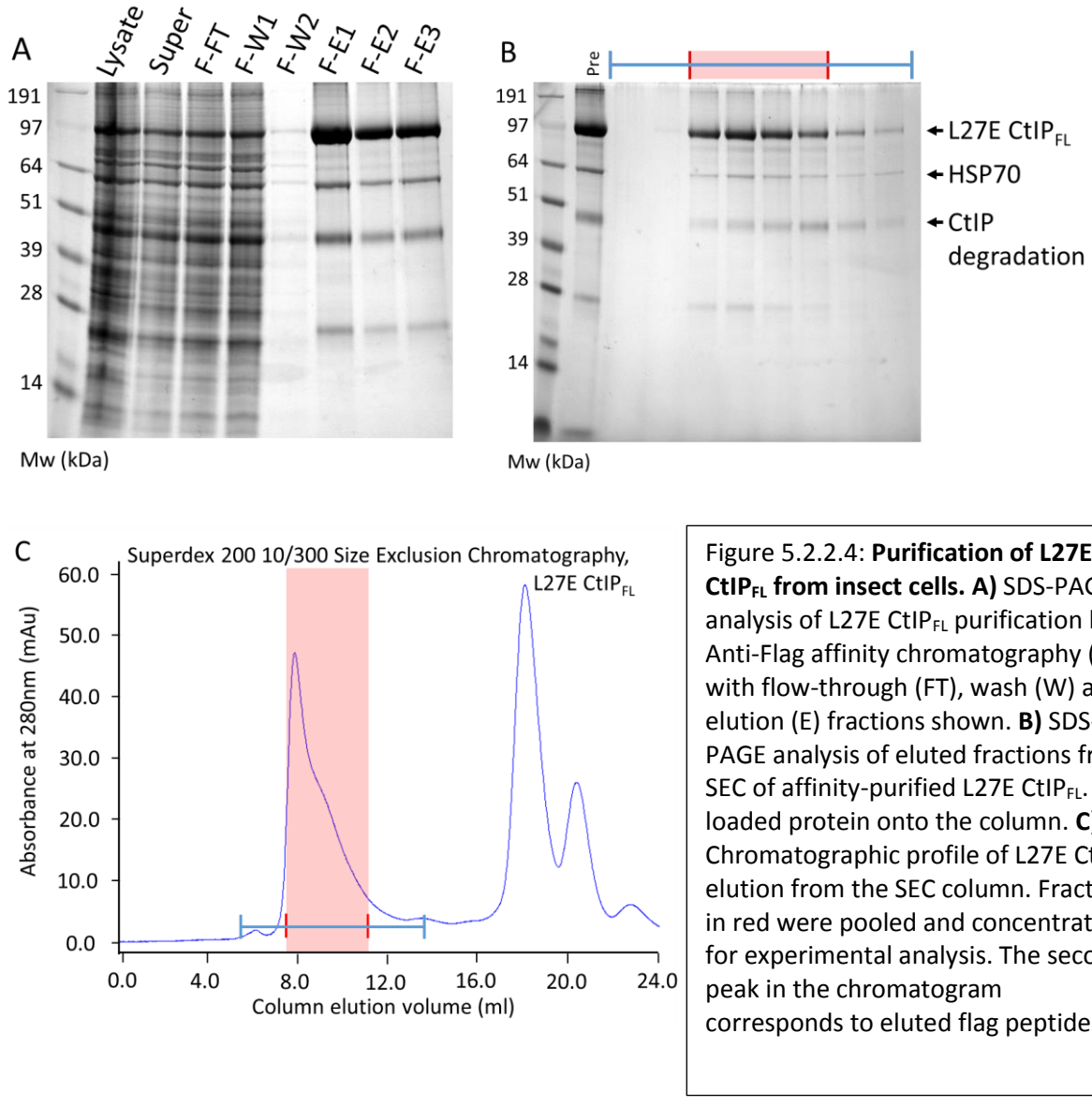


Figure 5.2.2.4: **Purification of L27E CtIP$_{FL}$ from insect cells. A)** SDS-PAGE analysis of L27E CtIP$_{FL}$ purification by Anti-Flag affinity chromatography (F), with flow-through (FT), wash (W) and elution (E) fractions shown. **B)** SDS-PAGE analysis of eluted fractions from SEC of affinity-purified L27E CtIP$_{FL}$. Pre: loaded protein onto the column. **C)** Chromatographic profile of L27E CtIP$_{FL}$ elution from the SEC column. Fractions in red were pooled and concentrated for experimental analysis. The second peak in the chromatogram corresponds to eluted flag peptide.

The CtIP$_{L27E}$ protein eluted in a wider peak and at a later volume (8.0 mL) than the wild-type protein (7.2 mL), in agreement with the expected reduction in oligomerisation status. The L27E CtIP$_{FL}$ sample achieved a similar degree of purity as the wild-type protein and was concentrated to 0.7 mg/mL or 6.75 µM (monomer).

5.2.2.5 CtIP$_{L27E}$ DNA binding

The EMSA was used for side-by-side comparison of L27E CtIP$_{FL}$ and the CtIP-CTD, using the same assay as in the previous EMSAs of CtIP$_{FL}$. The DNA substrate was a 200 bp fluorescently-labelled dsDNA at a final concentration of 1 µM.

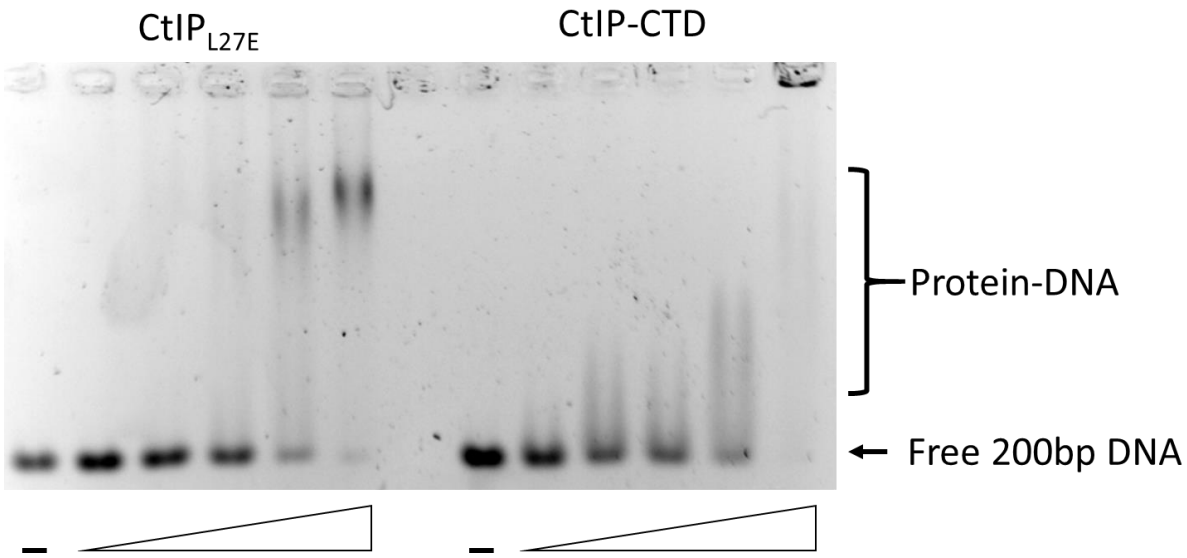


Figure 5.2.2.5: **DNA binding of L27E CtIP$_{FL}$ by EMSA.** Left-hand side - EMSA of 1 µM 200 bp dsDNA in the presence of L27E CtIP$_{FL}$ at concentrations of 6 µM, 3 µM, 1.5 µM, 0.75 µM, and 0.375 µM. Right-hand side: EMSA of 1 µM 200 bp dsDNA in the presence of CtIP-CTD at concentrations of 12 µM, 6 µM, 3 µM, 1.5 µM and 0.75 µM. The position in the gel of the free dsDNA and the slower migrating protein-DNA complexes are indicated.

The EMSA shows that the L27E CtIP$_{FL}$-DNA complex shows a similar electrophoretic behaviour to that of CtIP$_{FL}$ (figure 5.2.2.5). Interestingly, in this experiment CtIP$_{L27E}$ achieves full bandshift of the DNA probe, suggesting a higher affinity for dsDNA than the wild-type protein. This might be due to the differences in overestimation of protein concentration between CtIP$_{FL}$ and L27E CtIP$_{FL}$. The observed binding of CtIP$_{L27E}$ to DNA indicates that tetramerisation is not necessary for interaction with DNA.

5.2.2.6 Expression and purification of Mre11-Rad50-Nbs1 (MRN)

The direct interaction of CtIP with the MRN complex is necessary for DNA-end resection in preparation for homologous recombination (Anand et al., 2016; San Filippo et al., 2008; Sartori et al., 2007). To explore this critical interaction, I set out to prepare purified MRN. The MRN complex was expressed and purified from *Sf9* cells infected with individual Mre11, Rad50 and Nbs1 viruses. The complex was initially purified from filtered cell lysate supernatant by fast-flow heparin chromatography, followed by Anti-Flag affinity chromatography to capture the complex using Mre11's Flag tag and a final step of Hi-Trap S ion-exchange chromatography (figure 5.2.2.6a). The identity of the purified MRN components was confirmed by MALDI-TOF MS to be Rad50 (155 kDa), Nbs1 (85 kDa) and Mre11 (81 kDa). Analysis of the protein by SDS-PAGE showed stoichiometric quantities of Mre11 and Rad50 however sub-stoichiometric levels of Nbs1. Nevertheless the purified protein was shown to possess manganese-dependent nuclease activity on a PhiX174 circular single-stranded virion DNA substrate using the published nuclease activity assay (Sartori et al., 2007) (figure 5.2.2.5b).

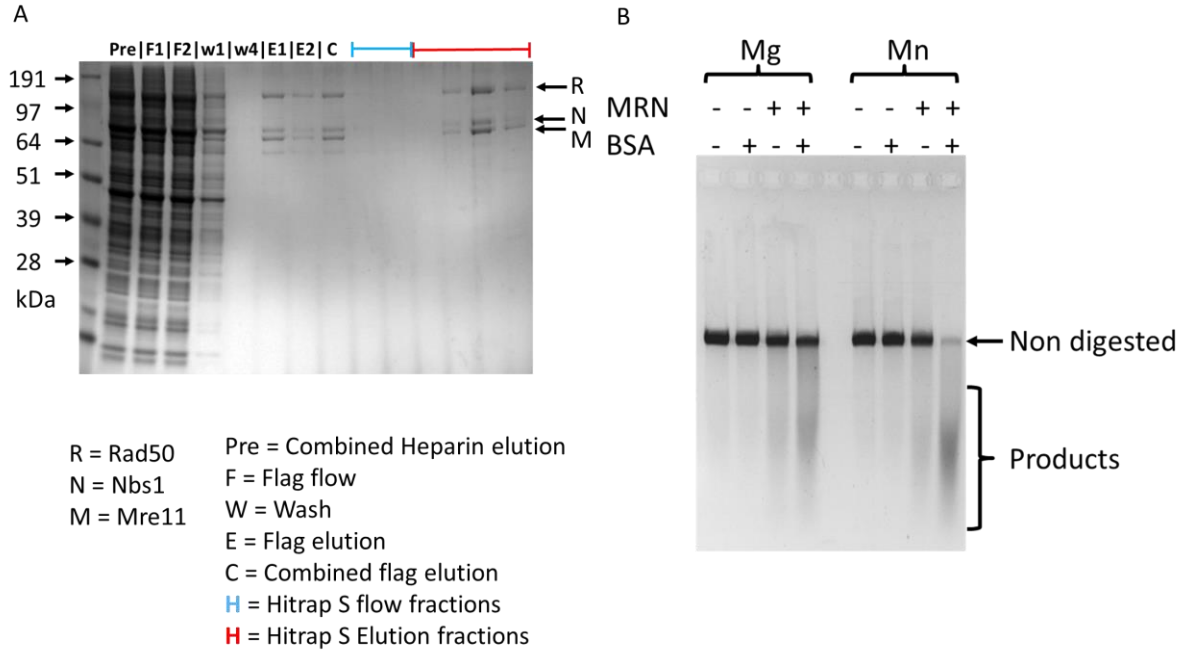


Figure 5.2.2.6: **Purification of Mre11-Rad50-Nbs1 from insect cells. A)** SDS-PAGE analysis of MRN purification by Anti-Flag affinity chromatography. **B)** Nuclease gel assay of MRN performed in accordance to Satori *et al* 2007 using purified protein from A. The assay was conducted with 40ng of MRN, 500ng of PhiX174 circular single-stranded virion DNA substrate and 5 mM $MgCl_2$ or 5 mM $MnCl_2$.

5.2.2.7 $CtIP_{FL}$ Protein purification optimisation

At this point a change of approach was necessary to increase initial capture and gain sufficient protein for analysis. For this, the purification tag was changed to the twin-Strep tag sequence to allow purification using the Strep-Tactin resin. This system was chosen as it has a greater binding capacity per mL along with the ability to use more stringent conditions to remove contamination. This would also allow pull-down experiments to be performed with the MRN complex.

5.2.2.7.1 Cloning of twin-Strep CtIP

The sequence corresponding to Human CtIP was PCR-amplified from the original flag tagged $CtIP_{FL}$ construct variant x3 and subcloned into both pACEBac1 for insect cell expression and pACEMam1 for mammalian expression. Positive clones were identified through restriction digest and DNA sequencing. Further insertion of the Twin-Strep tag was achieved to generate an N-terminally Strep tagged $CtIP_{FL}$. Clones were tested by *Bst*BI cleavage that had been engineered into the tag, followed by DNA sequencing to confirm the correct orientation of the tag.

5.2.2.7.2 Twin Strep $CtIP_{FL}$ initial purification.

One litre of *Sf9* cells were infected with the new virus and harvested at +2 days. Soluble protein was diluted to 100 mM NaCl for heparin binding and filtered before application to the column via a sample pump. Protein was eluted in a sharp peak with 70% v/v buffer B (1 M NaCl). This 5 mL elution was subsequently bound to a 1 mL Hi-Trap Strep-Tactin column, washed, eluted and fractions analysed by SDS-PAGE at each stage. Western blot analysis was used to determine possible effects on protein solubility of buffer salt concentration and the efficacy of the heparin and Strep-Tactin purification steps (figure 5.2.2.7.2). 100 mM NaCl had little apparent effect on CtIP solubility, however a large amount of CtIP protein could be detected within the heparin flow through (Hep-FT). Western blot of the Strep elution only contains $CtIP_{FL}$, conversely on SDS-PAGE analysis (figure 5.2.2.7.2b) many degradation products are seen indicating not all co-purification of CtIP degradation products contain the Strep-tag.

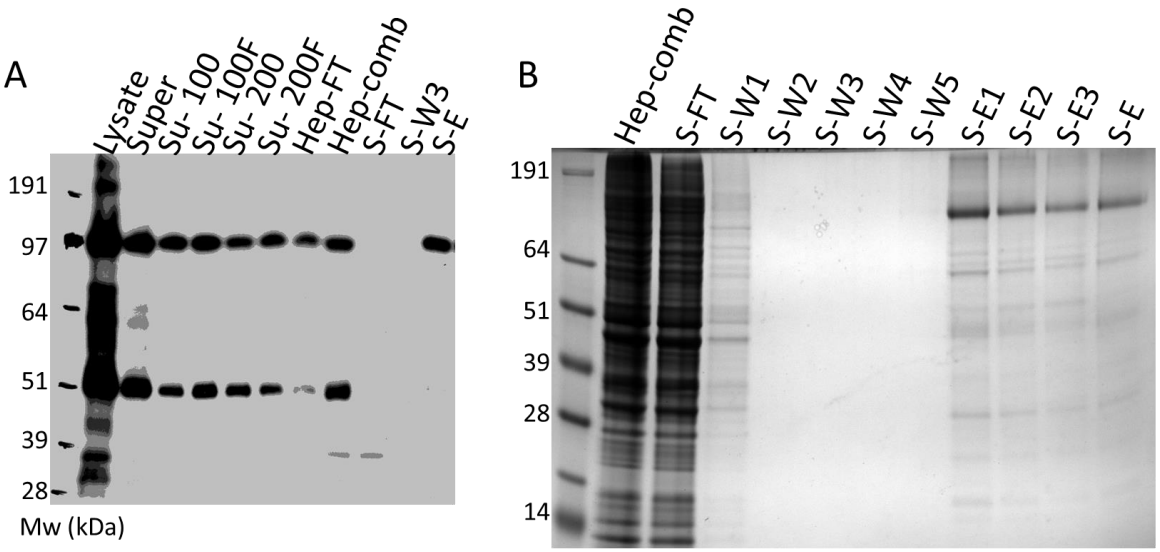


Figure 5.2.2.7.2: **Purification of $CtIP_{FL}$ twin strep tagged protein using Hi-Trap heparin and Strep-Tactin resin. A)** Western blot analysis of initial purification stages shows little change in protein solubility with 300 mM (super), 100 mM (su-100) or 200 mM (su-200) NaCl concentrations. Over both images Hi-Trap heparin (hep) and Strep-Tactin (S) with flow-through (FT), wash (W) and elution (E) purification fractions are shown. Little $CtIP_{FL}$ is detected in the Strep flow through with only $CtIP_{FL}$ detected by Western Blot **B)** SDS-PAGE analysis of Strep-Tactin capture and elution.

5.2.2.7.3 Twin Strep $CtIP_{FL}$ ammonium sulphate precipitation

Ammonium sulphate precipitation is a classical method for initial purification of proteins by salting out (Wiley, 1998). To determine if this method could remove CtIP degradation products, powdered ammonium sulphate was added to filtered cell lysate supernatant and precipitated protein removed by centrifugation. Fractions were dialysed into standard buffer to allow analysis of retained $CtIP_{FL}$ at each stage by SDS-PAGE and Western blot using the anti-Strep antibody (figure 5.2.2.7.3). $CtIP_{FL}$ protein was detected within the 20% w/v ammonium sulphate fraction and was impossible to be further purified by Strep-Tactin capture using the standard procedure and buffers.

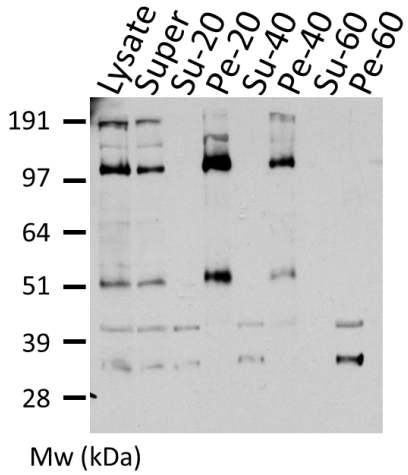


Figure 5.2.2.7.3: **Anti-strep Western blot analysis of ammonium sulphate precipitation fractions.** Fractions are sequential from a one litre insect cell purification of twin-strep $CtIP_{FL}$. "Su" and "super" stand for supernatant, "Pe" stands for the insoluble proteins removed by centrifugation with numbers indicating the percentage of ammonium sulphate. Fractions were dialysed into standard buffer overnight to allow SDS-PAGE and Western blot analysis. $CtIP_{FL}$ and its degradation products were detected within the insoluble pellet fractions.

5.2.2.7.4 Direct capture of Twin Strep CtIP$_{FL}$

Two litres of *Sf9* cells infected with twin-Strep CtIP$_{FL}$ baculovirus were gown and harvested at +2 days. Benzonase® was added to half of the cell preparation with filtered cell lysate supernatant passed over two individual 1 mL Strep-Tactin columns. It was apparent that the Benzonase® treated sample flowed through the column faster by gravity flow. The column was washed with the retained CtIP$_{FL}$ protein eluted with *d*-desthiobiotin elution buffer. Fractions were analysed by SDS-PAGE with a considerable amount of degradation products co-eluting (figure 5.2.2.7.4).

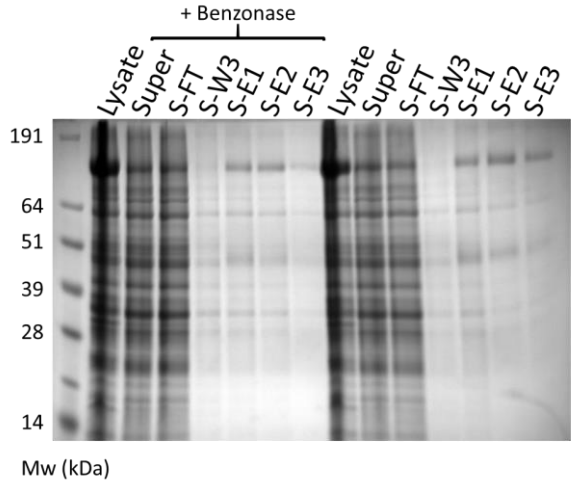


Figure 5.2.2.7.4: **Insect expressed CtIP$_{FL}$ strep capture with and without Benzonase®.** Protein was expressed using the baculovirus expression system with the preparation split into two and purified identically. The Strep-Tactin resin (S), Flow-through (FT), Wash (W) and Elution (E) are shown.

Although CtIP$_{FL}$ was purified using the insect cell baculovirus system the protein was unstable with an increase in CtIP$_{FL}$ degradation products detected by SDS-PAGE. This instability is likely due to the unknown complexity of the human CtIP unstructured region. Thus, an alternative human expression system was chosen for further expression of twin-Strep CtIP$_{FL}$.

5.2.3 Human Embryonic Kidney 293F expressed Full Length CtIP

Production of recombinant $CtIP_{FL}$ was changed to the Human Embryonic Kidney 293F suspension cell line to increase the likelihood that the protein was active, stable, correctly modified, and true to the real human protein; this should then give the most reliable results. Maintenance of the HEK293F cell line and tissue culture facility was performed by Dr Neil Rzechorzek and Vincentius Aji Jatikusumo.

5.2.3.1 Protein purification of $CtIP_{FL}$ optimisation

A small-scale expression test was performed infecting 50ml of HEK293F cells at $1X10^6$ per mL with the CtIP Twin Strep pACEmam1 expression construct. At three time intervals (+1 day, +2 days, +3 days) 1 mL of cells were removed for expression testing. Cells were pelleted and suspended in lysis buffer: 20 mM Tris pH 8, 300 mM NaCl, 0.5% v/v Tween 20, 10% v/v glycerol and protease inhibitor tablets (EDTA-Free). Total lysate and supernatant of each time point were analysed by Western blot using an Anti-Strep antibody to detect the Strep-tagged $CtIP_{FL}$ protein. Expression of $CtIP_{FL}$ was detected from +1 days onwards, together with evidence of degradation products (figure 5.2.3.1-1a). From this experiment it was therefore determined that cells should be harvested at +2 days to ensure a maximum $CtIP_{FL}$ yield and minimise the formation of degradation products. 400 mL of HEK293F cells at $1X10^6$ per mL with the CtIP Twin Strep pACEmam1 expression construct was harvested at +2 days using the same procedure however protein was absent from the soluble fraction by anti-Strep Western blot (figure 5.2.3.1-1b).

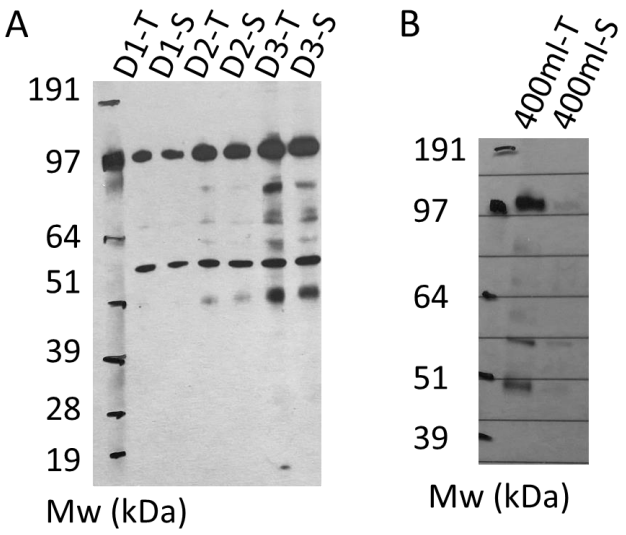


Figure 5.2.3.1-1: **Western blot of 50 mL and 400 mL $CtIP_{FL}$ HEK293F expression tests using Anti-Strep antibody.** "D" stands for day, "T" for total lysate and "S" for clarified lysate supernatant for both images. **A)** Time trial expression test with 50ml of cells. Day2 (D2) was chosen for subsequent harvesting of HEK293F cells in B. **B)** 400 mL expression of $CtIP_{FL}$ in HEK293F cells. Most of the protein was shown to be insoluble.

A further 400 mL of HEK293F cells expressing the $CtIP_{FL}$ construct were prepared, however, 0.5 mM TCEP was added to the cell re-suspension buffer. This was shown to improve solubility of CtIP by Anti-Strep Western blot analysis. Filtered cell lysate supernatant was applied to a 1 mL Strep-Tactin column, washed, eluted and fractions analysed by SDS-PAGE (figure 5.2.3.1-2). $CtIP_{FL}$ was detected within the elution fractions by SDS-PAGE, however with insufficient yield for biochemical analysis.

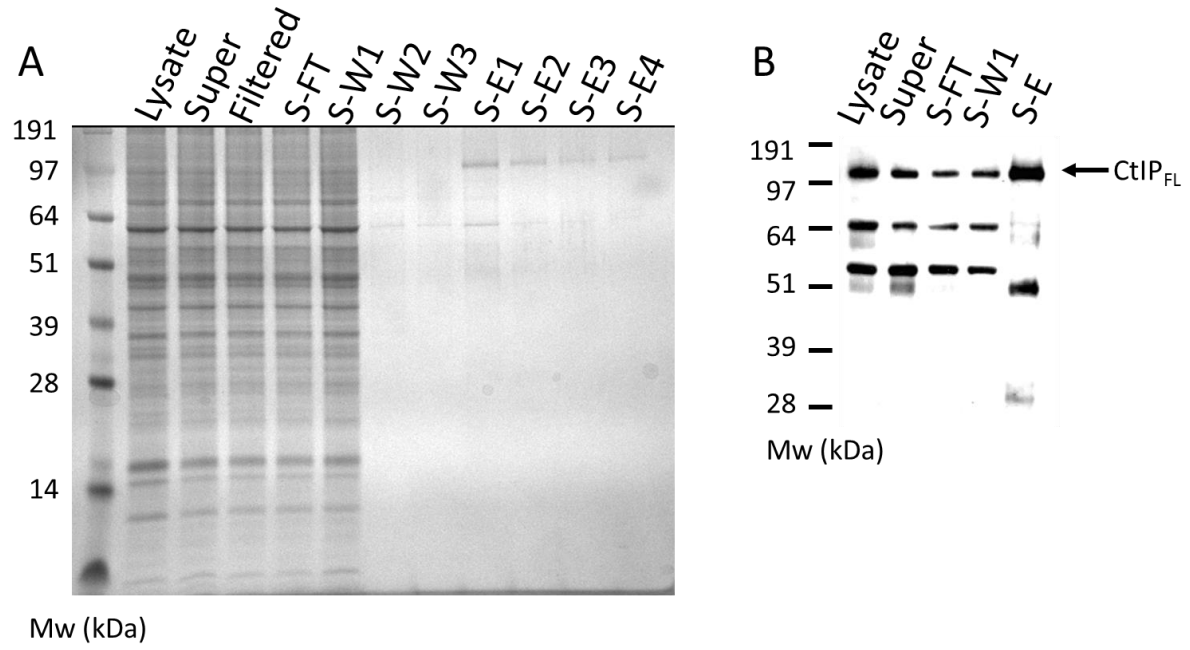


Figure 5.2.3.1-2: **HEK293F expressed $CtIP_{FL}$ purification using Strep-Tactin resin. A)** SDS-PAGE analysis of $CtIP_{FL}$ purification by Strep-Tactin chromatography (S) with flow-through (FT), wash (W) and elution (E) fractions shown. **B)** Anti-Strep-tag Western blot, $CtIP_{FL}$ can be detected in all fractions.

It has been recently reported that addition of valproic acid can enhance recombinant protein production in transiently infected cell lines (Backliwal et al., 2008; Wulhfard et al., 2010). A further expression time trial was repeated with final concentration 4 mM valproic acid added to the cell media. The same procedure as detailed previously was performed with 1 mL samples taken at one day intervals until +5 days. Total lysate and supernatant were analysed by anti-Strep Western blot (figure 5.2.3.1-3).

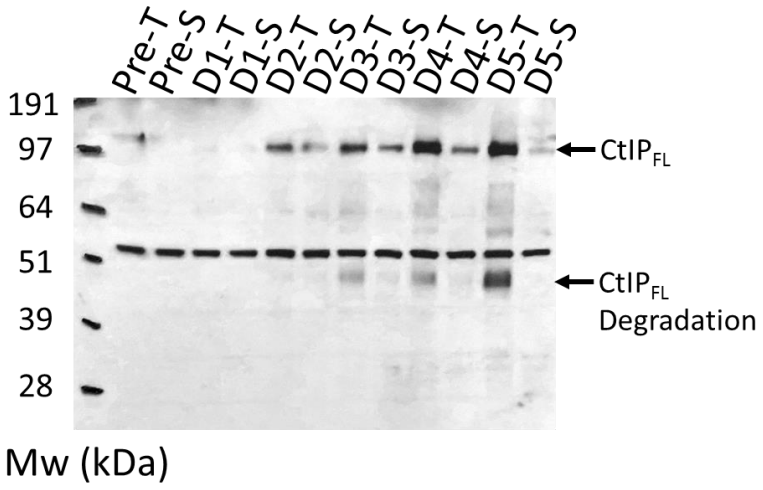


Figure 5.2.3.1-3: **Western blot of CtIP$_{FL}$ HEK293F valproic acid expression test using Anti-Strep antibody.** Samples corresponding to total protein (T) and supernatant (S) from clarified lysate are shown on the Western blot. 'Pre' were cells not transfected with the CtIP$_{FL}$ construct, the band at 51kDa is therefore from the HEK293F cells.

Addition of valproic acid increased soluble protein yield with optimal cell harvesting determined to be +4 days. HEK293F cells not transfected gave a contamination band above 51 kDa detectable by Anti-Strep Western blot. However, the band was not present on Strep-Tactin elution of CtIP$_{FL}$ so was no longer pursued (figure 5.2.3.1-3).

The addition of 0.5 mM TCEP on cell lysis improved solubility. Further optimisations were tested through small-scale analysis of valproic acid treated cells. Increasing Tris concentration to 50 mM had a greater yield of protein from Strep-Tactin elution with Benzonase® increasing the speed of gravity flow chromatography without reducing yield (figure 5.2.3.1-4).

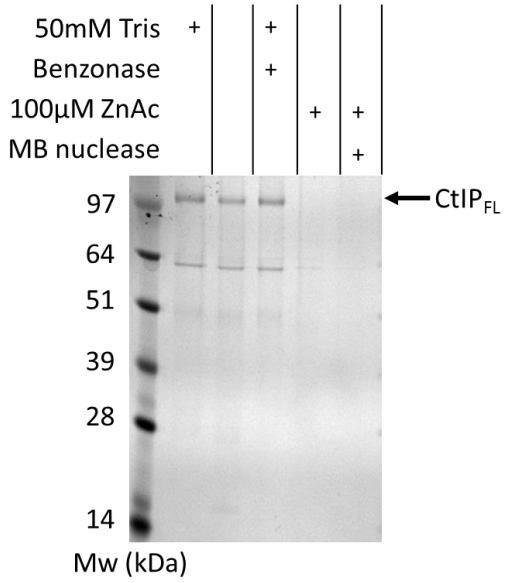
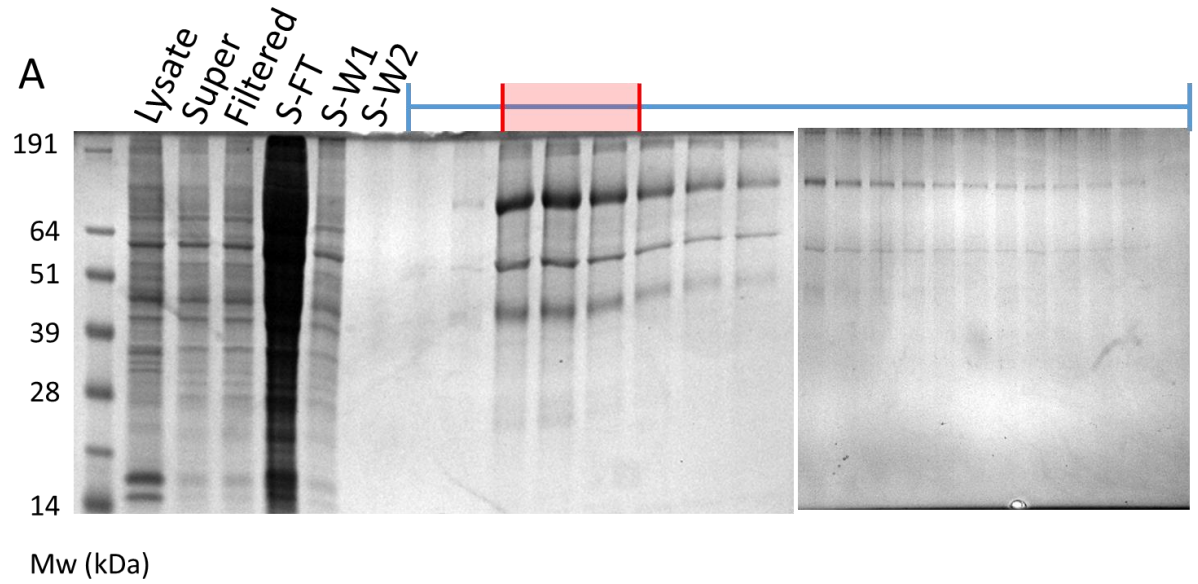
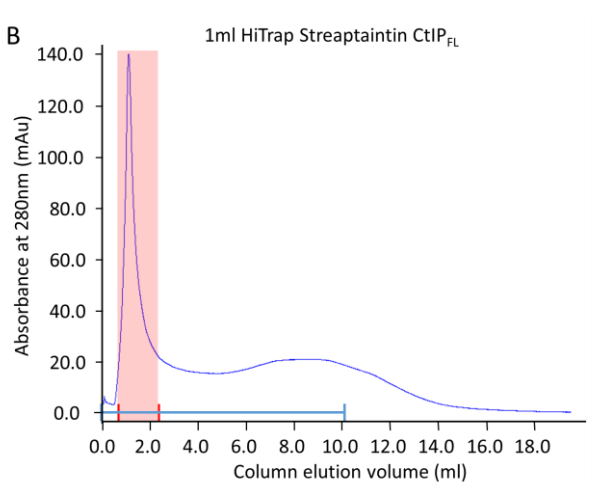
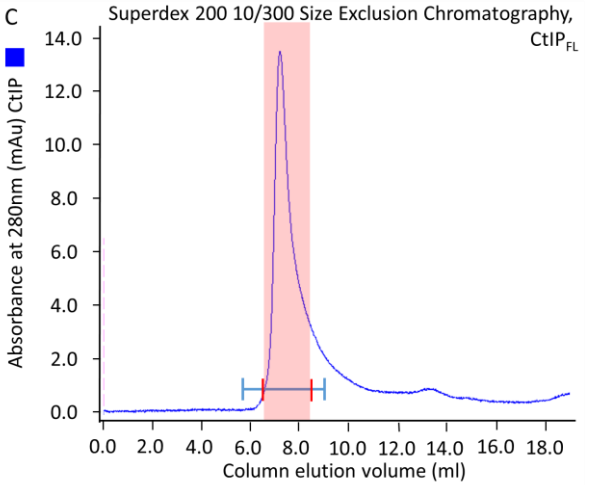


Figure 5.2.3.1-4: **50 mL HEK293F CtIP$_{FL}$ expression test with valproic acid and Strep-Tactin capture.** CtIP$_{FL}$ was detectable by Coomassie Brilliant blue stain on SDS-PAGE analysis indicated by the arrow. Lane 2 was a control preparation using 20 mM Tris and 100 mM NaCl. CtIP$_{FL}$ could not be detected in Strep-Tactin elution with addition of zinc or Mung Bean (MB) nuclease.

Purification optimisations were combined for a 400ml HEK293F expression test of Strep-tagged CtIP$_{FL}$ grown in the presence of 4mM valproic and harvested at +4 days. Cells were pelleted and suspended in lysis buffer: 50 mM Tris pH 8, 100 mM NaCl, 0.5 mM TCEP, 0.5% v/v Tween 20, 10% v/v glycerol, 2.5 units of Benzonase® and SigmaFAST™ EDTA-Free protease inhibitors. Filtered cell lysate supernatant was first passed over a 1 mL Hi-Trap Strep-Tactin column, washed and the retained CtIP$_{FL}$ eluted. Fractions were analysed by SDS-PAGE and pooled accordingly; the early eluting CtIP$_{FL}$ was applied in two 500 µL applications to a Superdex 200 10/300 SEC column pre-equilibrated into buffer containing 50 mM Tris pH 8, 300 mM NaCl, 0.5mM TCEP and 10% v/v glycerol. Eluted fractions containing CtIP$_{FL}$ as shown by SDS-PAGE were pooled and concentrated to 0.52 mg/mL 5.6 µM in 500 µL (figure 5.2.3.1-5.







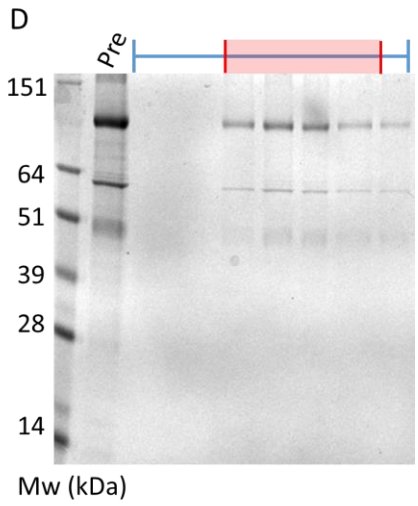


Figure 5.2.3.1-5: **HEK293F expressed CtIP$_{FL}$ Purification using Strep-Tactin resin and SEC. A)** SDS-PAGE analysis of CtIP$_{FL}$ purification by Strep-Tactin affinity chromatography (S) with flow-through (FT), wash (W) and elution fractions corresponding to the chromatogram in B. **B)** Chromatographic profile of CtIP$_{FL}$ elution from the Strep-Tactin column. Fractions highlighted in red were pooled for SEC. **C)** Superdex 200 10/300 SEC of CtIP$_{FL}$, two runs were performed with one shown. Fractions from run 1 were analysed by SDS-PAGE in D. **D)** SDS-PAGE analysis of eluted fractions from size-exclusion chromatography of affinity-purified CtIP$_{FL}$. Pre: protein loaded onto the column. Fractions highlighted in red pooled and concentrated for analysis.

5.2.3.2 Dextran Blue void volume calculation for the Superdex 200 10/300 SEC column.

$CtIP_{FL}$ has been shown to elute from SEC with an estimated size of 766 kDa, almost double its molecular weight (Davies et al., 2015). As 766 kDa is above the gel matrix exclusion limit for a Superdex 200 SEC column the volume of buffer surrounding the beads, termed void volume, was determined. Blue dextran has a weight of 2 mega Daltons so cannot flow through the gel matrix, its elution volume can therefore be interpreted as the void volume. Figure 5.2.3.2 shows the two SEC chromatograms overlaid with the blue dextran peak (7.1 mL) eluting earlier than $CtIP_{FL}$ (7.6 mL). Thus, showing the bulk of $CtIP_{FL}$ is flowing through the gel matrix rather than the void of the column. The $CtIP_{FL}$ peak is however broad and overlaps with blue dextran indicating there is a mixture of species with some too large to pass through the gel matrix.

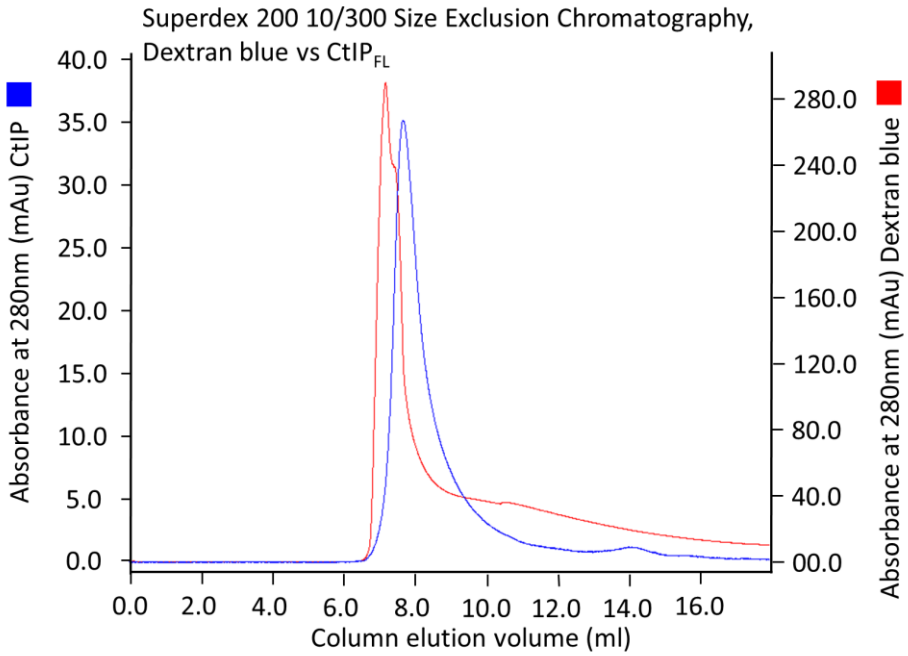


Figure 5.2.3.2: **Superdex 200 10/300 SEC of $CtIP_{FL}$ (blue) overlaid with Dextran blue (red).** Y-axis correspond to the two peaks and have been adjusted to show full peak comparison.

5.2.3.3 DNA binding

The EMSA was further used to assess the DNA binding affinity of HEK293F expressed CtIP. A 60bp fluorescently-labelled dsDNA or ssDNA substrate at a final concentration of 100 nM was used with $CtIP_{FL}$ protein concentration from 5 µM to 0.31 µM. The addition of CtIP results in formation of a slow migrating species representing a protein-DNA complex. Increasing concentrations shift the equilibrium from free-DNA to bound-DNA indicating further protein binding (figure 5.2.3.3-1).

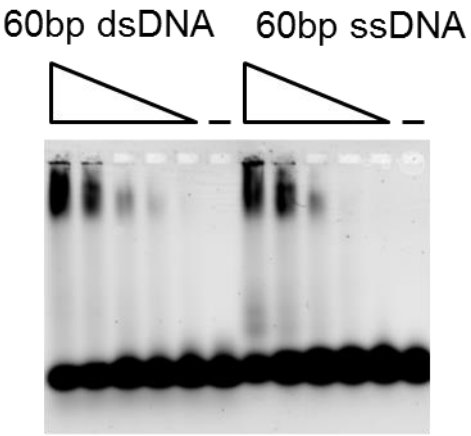


Figure 5.2.3.3-1: **Binding of $CtIP_{FL}$ to 60 bp dsDNA and ssDNA by EMSA.** Triangle represents decreasing estimated concentrations of CtIP from 5.0 µM to 0.31 µM with 100 nM DNA. No major differences seen between dsDNA and ssDNA.

To deduce the minimum length of DNA necessary for the CtIP-DNA interaction, further EMSAs were performed using the same procedure with progressively shorter dsDNA substrate. Initial experiments showed DNA binding to 40 bp dsDNA with minimal signs of DNA binding for substrates of 20 bp dsDNA and below. The length of substrates was reduced by 5 bp dsDNA iteratively with similar amounts of binding shown from 40 bp to 25 bp dsDNA. The DNA binding of CtIP to 20 bp is greatly impaired with no binding at low concentrations and a large reduction at 5.0 µM. These results indicate the minimum length requirement of DNA for CtIP binding to be between 25 bp and 20 bp (figure 5.2.3.3-2). This is the same as previously determined for CtIP-CTD (Chapter 4).

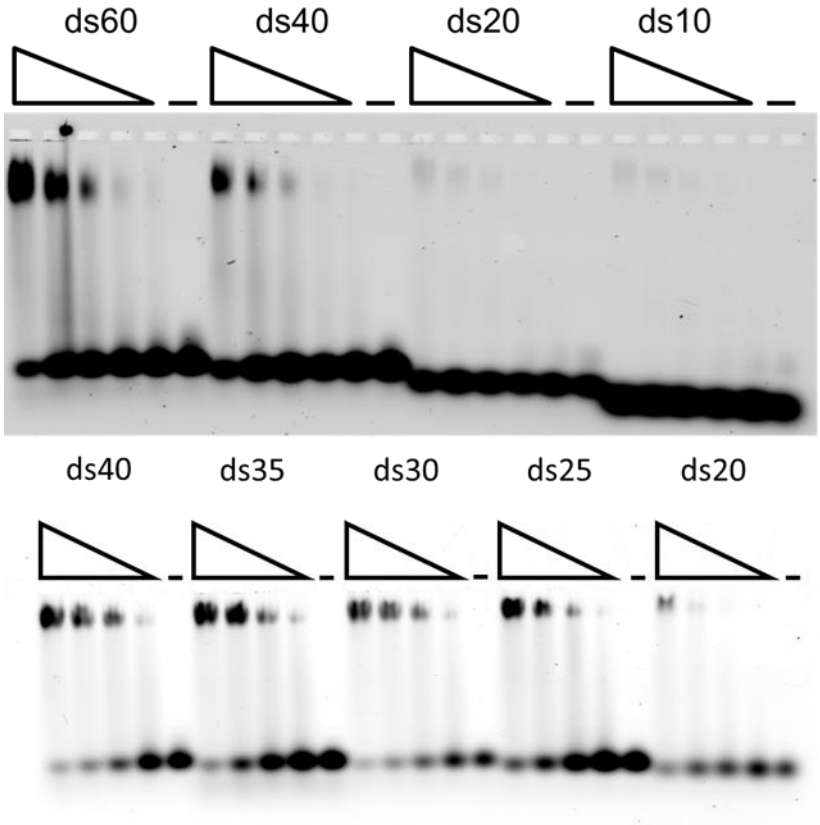


Figure 5.2.3.3-2: **Determination of minimum DNA length for $CtIP_{FL}$ interaction by EMSA.** Top panel- triangle represents decreasing estimated concentrations of CtIP from 5.0 µM down to 0.31 µM. Left to right: EMSAs with 60 bp, 40 bp, 20 bp and 10 bp fluorescently labelled 1 µM dsDNA. Bottom panel – triangle represents decreasing CtIP concentrations from 5.0µM to 0.62 µM. left to right, 40 bp, 35 bp, 30 bp, 25 bp and 20 bp 100 nM fluorescently labelled dsDNA. Addition of CtIP results in the formation of a slower migrating species which represents protein-DNA complexes.

The CtIP interaction with MRN has been reported to generate a nick 5' to the DSB on the 3' DNA strand (Anand et al., 2016). To test if this nick changes the DNA binding affinity of CtIP, 60 bp substrates were generated with a nick or 2 bp deletion on either the 3' or 5' strand. No differences were seen between the four different substrates each producing a similar amount of slower migrating DNA-CtIP complexes (figure 5.2.3.3-3). Conversely, there was a greater CtIP-DNA affinity for the linear 60 bp dsDNA compared to the nicked or 2 bp deleted substrates.

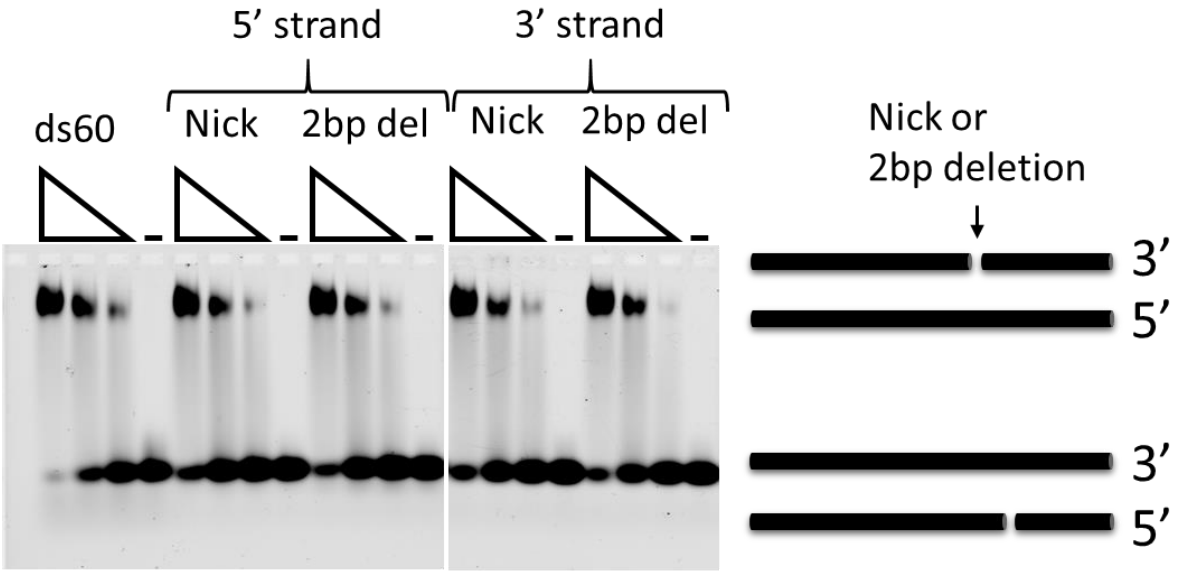


Figure 5.2.3.3-3**: Interaction of $CtIP_{FL}$ with nicked or 2bp deleted fluorescent DNA by EMSA.** Triangle represents decreasing estimated $CtIP_{FL}$ concentrations from 2.5 µM to 0.62 µM. Left to right: 60 bp fluorescently labelled dsDNA, 60 bp fluorescently-labelled dsDNA with 5' strand nick, 5' strand 2 bp deletion, 3' strand nick, 3' strand 2 bp deletion. Addition of CtIP results in the formation of a slower migrating species which represents protein-DNA complexes. No major differences were seen in relation to each different substrate over the three repeats. All DNA substrates were used at a final concentration 100 nM.

Resection of a DNA DSB by MRN and CtIP results in a 3’ DNA tail, necessary for homology directed DSB repair. To test if CtIP can bind to a single stranded DNA adjacent to a double stranded DNA, four further DNA substrates were generated. As the minimum interaction length was determined to be around 20 bp, two different lengths of ssDNA were generated: 24 bp and 16 bp. No differences were seen between the four oligoes with the same slight reduction in binding compared to 60 bp dsDNA (figure 5.2.3.3-4).

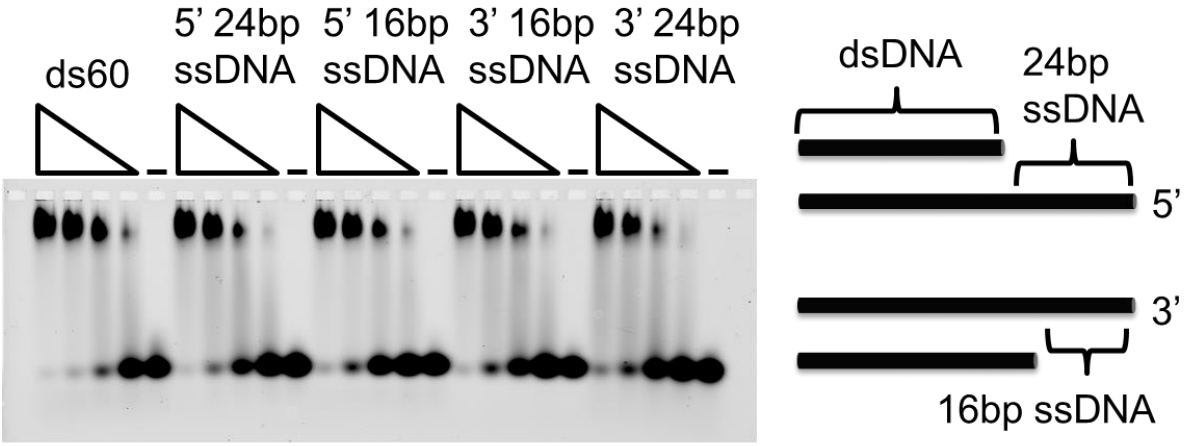


Figure 5.2.3.3-4**: Interaction of CtIP with 5’ or 3’ ssDNA tails on fluorescent DNA by EMSA.** Triangle represents decreasing estimated CtIP concentrations from 5.0 µM to 0.62 µM. From left to right: 60 bp fluorescently labelled ds DNA, 60 bp fluorescently labelled DNA: with 5’ strand 24 bp ssDNA, 5’ strand 16 bp ssDNA, 3’ strand 16 bp ssDNA, with 3’ strand 24 bp ssDNA. Substrates are represented on the right-hand side. Addition of CtIP results in the formation of a slower migrating species which represents protein-DNA complexes. No major differences were seen in relation to each different resected substrate over the three repeats. All DNA substrates were used at a final concentration 100 nM.

5.2.3.4 Nuclease activity

The nuclease activity of HEK293F-expressed $CtIP_{FL}$ was analysed using the protocol of Sartori *et al.* 2007. Either 200 or 350 ng of twin-Strep $CtIP_{FL}$ expressed in HEK293F cells was incubated with the PhiX174 substrate with either 5mM magnesium or manganese, analysed by gel electrophoresis and stained with Syber™ Gold. No nuclease activity was reported with either metal with identical results as addition of buffer (figure 5.2.3.4). The activity of HEK293F expressed CtIP with MRN could not be assessed due to lack of resources.

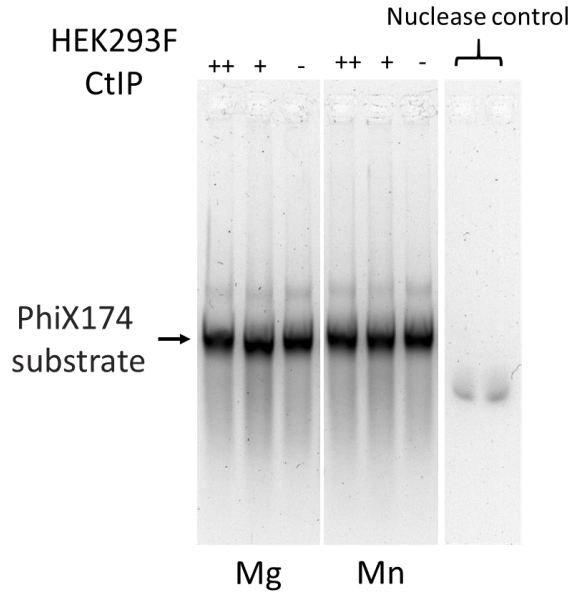


Figure 5.2.3.4**: HEK293F expressed $CtIP_{FL}$ does not have nuclease activity**. PhiX174 DNA substrate was incubated with either 350 ng (++) or 200 ng (+) of $CtIP_{FL}$ in 5 mM $MgCl_2$ or $MnCl_2$ run on an agarose gel and stained with SYBR Gold. Nuclease control was 200 ng of Benzonase with either Mg or Mn.

5.2.3.5 Measuring HEK293F expressed CtIP$_{FL}$ stability

HEK293F-expressed CtIP$_{FL}$ was observed to have a greater stability at room temperature compared to insect cell expressed CtIP$_{FL}$ protein (Data not shown). This is likely due to the differences in post translational modifications between the two expression systems, or the presence of contaminant proteases within the insect cell purification. To quantify the CtIP$_{FL}$ stability a thermal shift assay was performed with a melt curve produced and melt peak temperature of 42.5°C recorded (figure 5.2.3.5). The value obtained is an ensemble of CtIP$_{FL}$, its degradation products, and HSP70. As heat shock proteins are known to have high melt temperatures and remain stable at 50° (Boswell-Casteel et al., 2015). Therefore, CtIP must contribute to the value obtained, indicating the presence of folded CtIP$_{FL}$ protein.

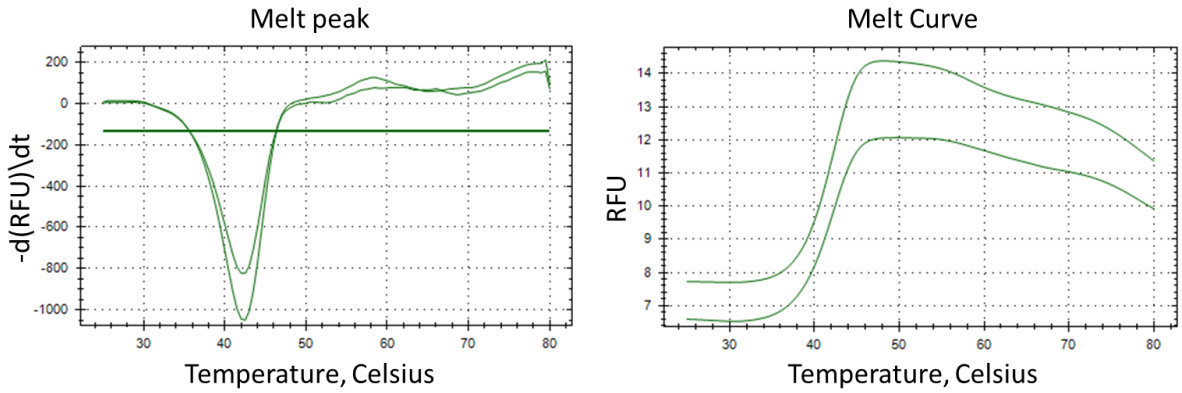


Figure 5.2.3.5: **Thermal melt peak and melt curve for HEK293F expressed CtIP$_{FL}$.** From the binding of SYPRO® Orange to unfolded protein the fluorescence increases and is measured in relative fluorescence units. Two concentrations of CtIP$_{FL}$ were used: 8.36 µM and 4.79 µM.

Further purification optimisation was performed with each iterative preparation of CtIP$_{FL}$ to increase yields. This led to a reverse flow elution from the Strep-Tactin column and increasing the volume of starting cell culture.

5.2.4 Cryo-EM of HEK293F Full Length CtIP

So far only small peptides and short domains of CtIP have been crystallised and the structure determined with CtIP predicted to have a large ~600 amino acid region of unstructured protein. Cryo-EM avoids the need for protein crystals but is typically restricted to large protein domains and complexes (>100 kDa). The Twin-Strep tagged CtIP has a molecular weight of 106.3 kDa, however following the CtIP tetramerisation the complex has a mass of 425 kDa which makes it a candidate for Cryo-EM to elucidate the structure of the full-length protein.

5.2.4.1 $CtIP_{FL}$ protein purification for Cryo-EM.

The $CtIP_{FL}$ Twin-Strep construct was expressed in 800 mL of HEK293F cells, with the filtered soluble protein applied to a 1 mL HiTrap Strep-Tactin column in three 10 mL applications. The column was washed, and flow reversed with an additional wash in reverse flow before the retained Strep-tagged $CtIP_{FL}$ was eluted. The reversal of the column decreased the elution volume and increased protein concentration (figure 5.2.4.1). Fractions were analysed by SDS-PAGE and pooled accordingly. 500 µL of protein was buffer exchanged using a Nap5 column to remove glycerol, with the peak 200 µL elution fraction at 0.9 mg/mL or 8.46 µM (monomer). A further 500 µL was applied to a Superdex 200 10/300 SEC column equilibrated into the standard buffer, the peak elution fraction gave a $CtIP_{FL}$ concentration of 0.24 mg/mL or 2.25µM (monomer). Both were used for preparation of Cryo-EM grids.

All grid manipulations were performed by Dr Neil Rzechorzek. The grids used were from Batch No.190698, QuantiFoil R 1.2/1.3 Cu300, see chapter 2 for details.

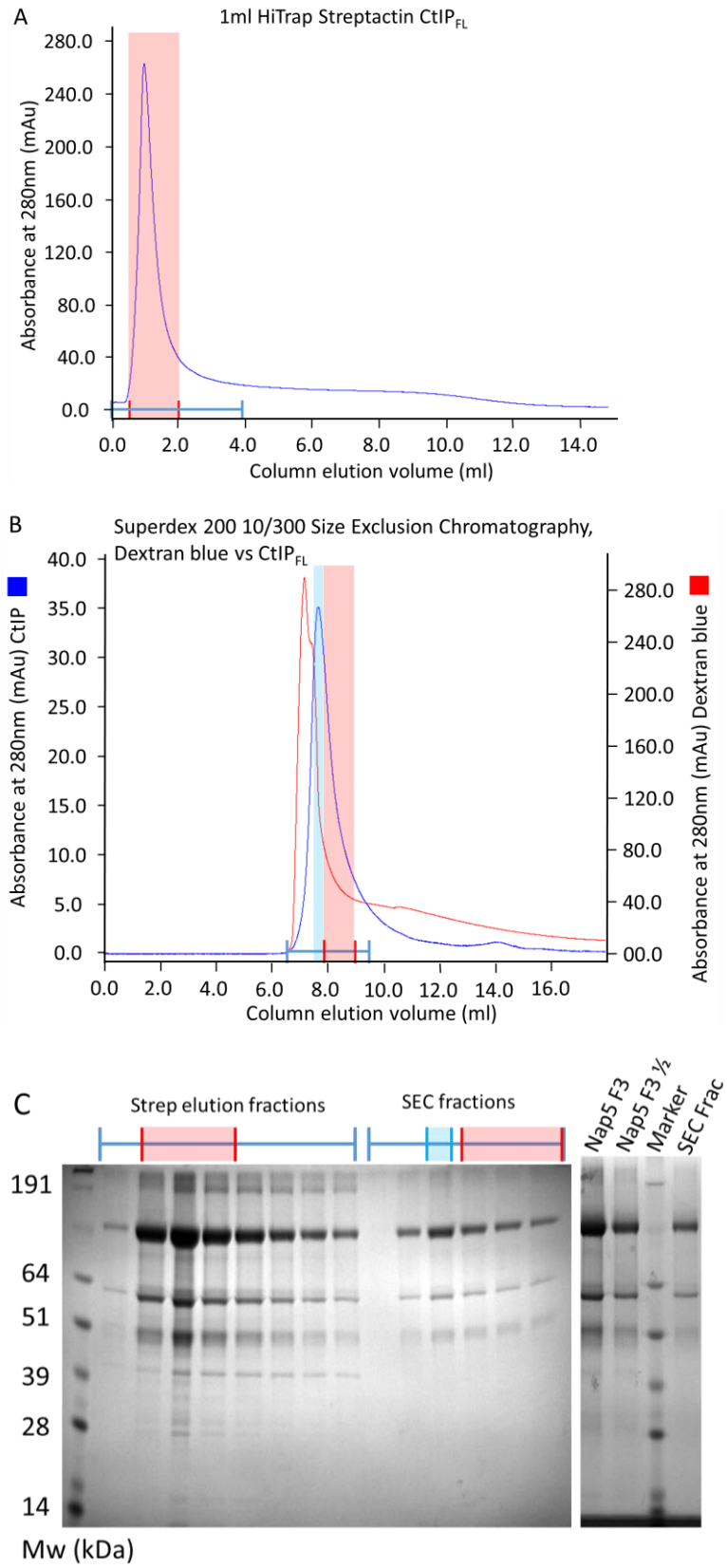


Figure 5.2.4.1: **HEK293F expressed CtIP$_{FL}$ Purification using Strep-Tactin and SEC for Cryo-EM.** CtIP protein purification from 800 mL of HEK293F cells expressing Twin-Strep tagged CtIP. **A)** Chromatographic profile of CtIP$_{FL}$ elution from the 1ml Hi-Trap Strep-Tactin column. Fractions in red were pooled with 500 µL used for SEC in B, 500 µL used for Nap5 buffer exchange with SDS-PAGE analysis of peak elution shown in C. **B)** Chromatographic profile of CtIP$_{FL}$ elution from the SEC column. Peak fraction in cyan was used for Cryo-EM Protein highlighted in red was pooled for further analysis. Red trace is Dextran Blue applied to the same column after CtIP$_{FL}$ had eluted. This is to calculate the void volume of the column and is detailed in section 5.2.3.2. **C)** SDS-PAGE analysis of fractions from the Strep-Tactin column and SEC. RHS of gel shows the fractions used for Cryo-EM. Nap5 F3 was the peak fraction from the Nap5 elution which was used at full and ½ concentration.

5.2.4.2 Cryo-EM Data collection

Grids containing $CtIP_{FL}$ were analysed using the Titan Krios G2 Transmission Electron Microscope (TEM), 300 keV, at the Nano Science department, University of Cambridge. The microscope was prepared by Dr Dima Chirgadze with the Falcon 3 detector on integration mode and controlled by Dr Jamie Blaza. It was noted that many squares per grid had dark black centres, indicating a non-optimal glow discharge. Large areas covered in thick ice proved problematic to image. Only a selection of squares had optimum ice thickness for image collection (figure 5.2.4.2-1).

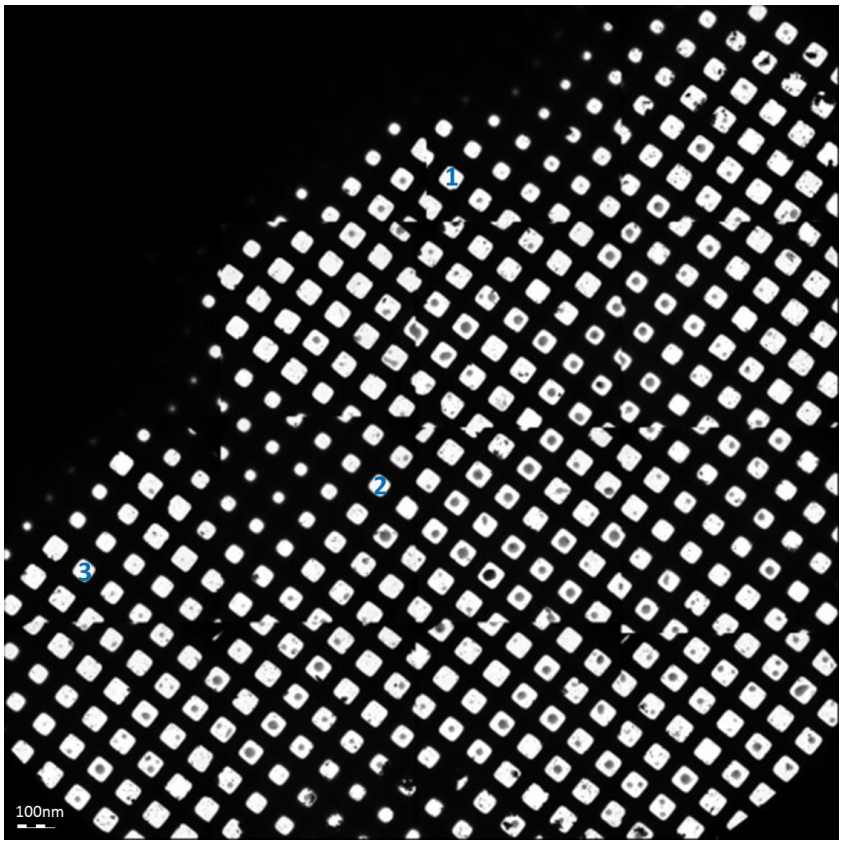


Figure 5.2.4.2-1: **Cryo-EM Atlas image from SEC purified protein.** Numbers indicate squares imaged.

Where imaging was possible protein appeared as dark dense with non-vitreous ice particles present across all micrographs. Across a series of 15 micrographs recorded 4 grid squares showed a recognisable structure which looked to be a filament. These filaments are shown in figure 5.2.4.2-3 with a representative complete micrograph shown in figure 5.2.4.2-2. Since the number of filaments per hole and the number of holes appropriate for screening were minimal further screening with optimised settings will be necessary. Filaments were not present in SEC filtered protein samples, likely because the protein was fractionated by size removing any mega Dalton complexes.

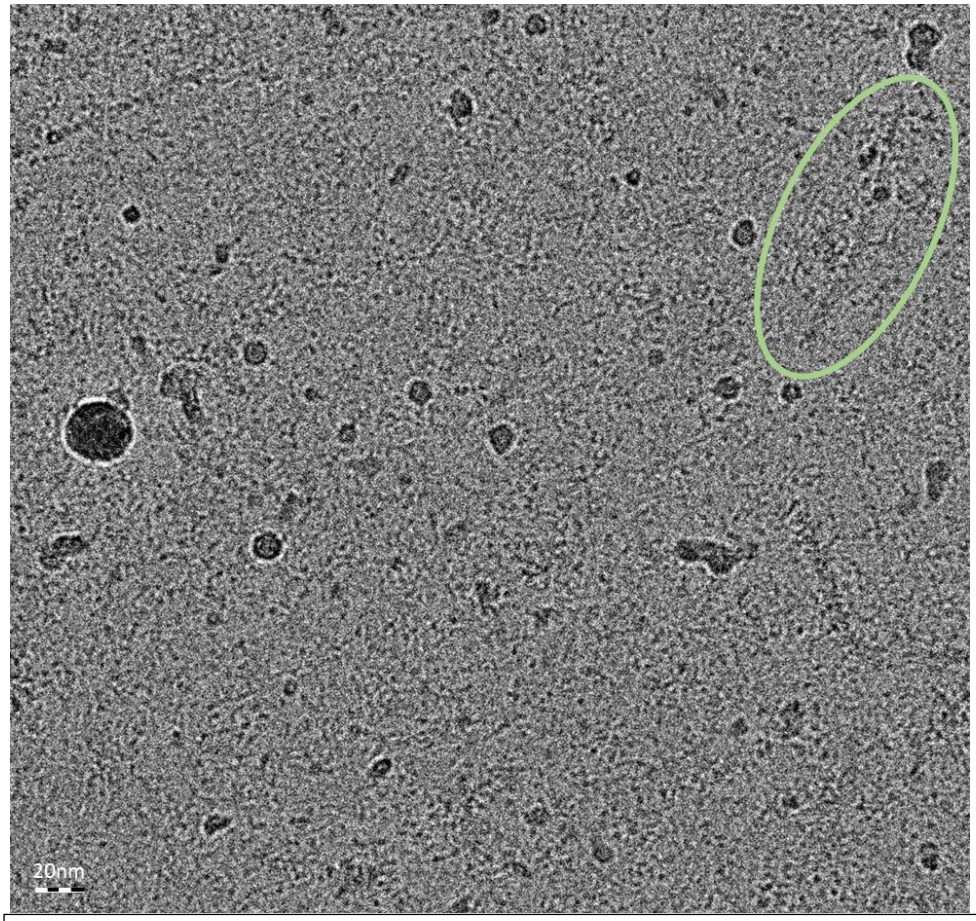


Figure 5.2.4.2-2: **CtIP$_{FL}$ Cryo-EM Titan Krios G2 microscope image.** Representative micrograph of Nap5 buffer exchanged protein applied to a QuantiFoil R 1.2/1.3 Cu300 grid. Non-vitreous ice can be seen as large dark circles. Various other particles can be seen, some of which are likely to be CtIP$_{FL}$ protein, such as filaments with an example in a green oval.

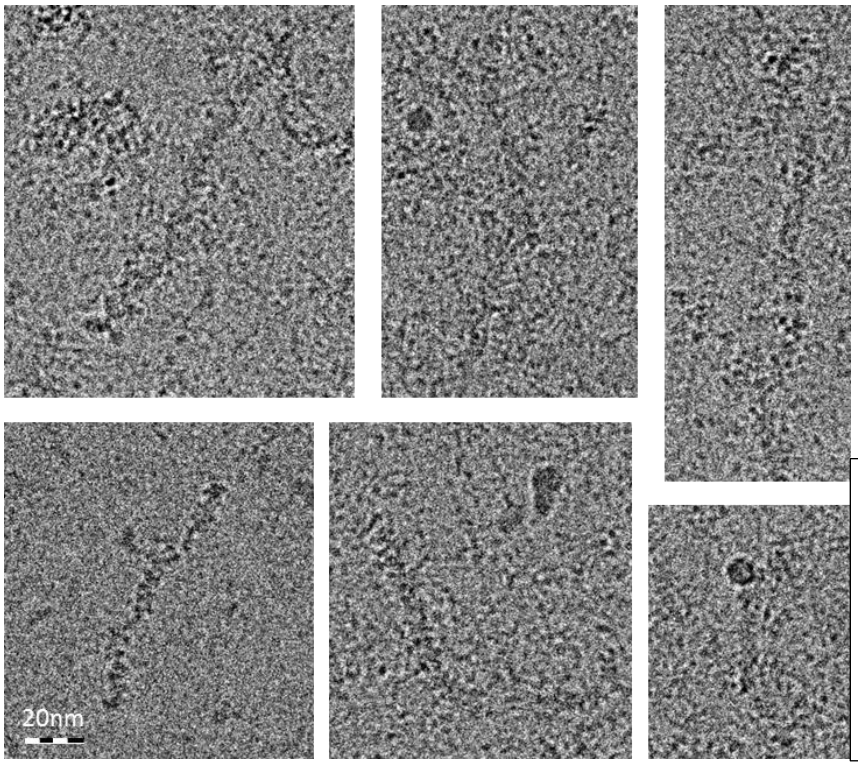


Figure 5.2.4.2-3: **CtIP$_{FL}$ Cryo-EM, filament image gallery.** Images span 4 grid squares from Nap5 buffer exchanged peak fraction protein.

5.2.4.3 CtIP Superose 6 SEC

$CtIP_{FL}$ protein elutes on a Superdex 200 close to the column exclusion limit and filaments have been seen in Cryo-EM micrographs. To separate the CtIP species a Superose 6 Increase SEC column was used. This column has a separation range for globular proteins of 5 kDa to 5 MDa and therefore should be able to separate different oligomers of CtIP. The $CtIP_{FL}$ Twin-Strep construct was expressed in 800 mL of HEK293F cells with the soluble protein filtered and applied to a 1 mL Hi-Trap Strep-Tactin column via Superloop. The column was washed with lysis buffer and flow reversed with an additional wash in reverse flow into the standard buffer (figure 5.2.4.3-1). Protein was eluted from the column with fractions containing $CtIP_{FL}$ concentrated to 10.52 mg/mL in 70 µL.

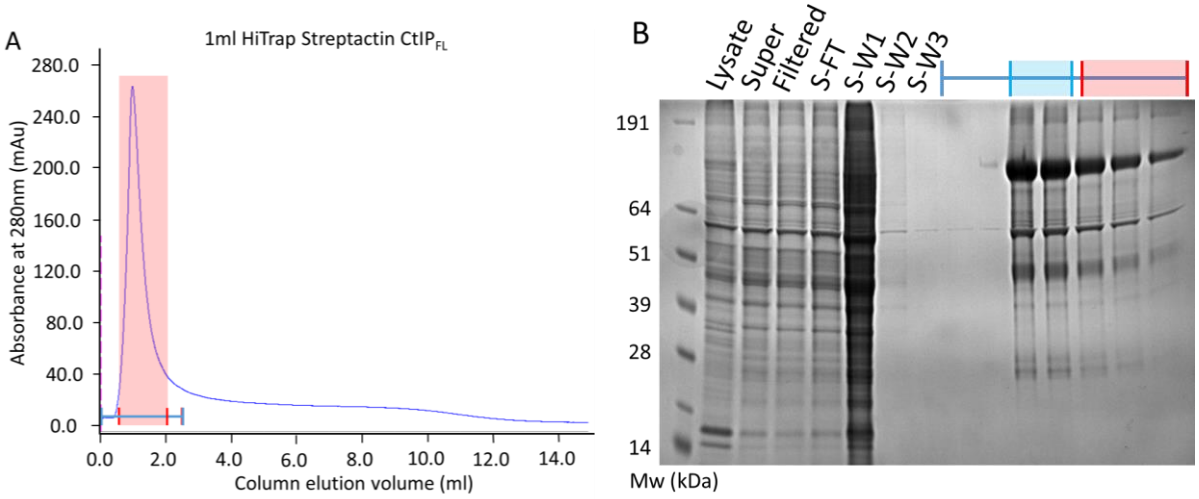


Figure 5.2.4.3-1: **Purification of $CtIP_{FL}$ from 800ml of HEK293F cells using Strep-Tactin resin. A)** Chromatographic profile of $CtIP_{FL}$ elution from the Strep-Tactin column. Fractions indicated are analysed by SDS-PAGE in B. Fractions in red were pooled for concentrating down to 70 µL for SEC analysis. **B)** Initial capture of protein by Strep-Tactin (S) resin with flow-through (FT) wash (W) and Elution fractions (E) shown.

The concentrated protein was centrifuged at 14,000 RCF 4°C with no change in concentration or appearance. To prepare the SEC column after equilibration sodium hydroxide was applied to clean the column. 60 µL of concentrated $CtIP_{FL}$ at 10.52 mg/mL was applied to the Superose 6 column and eluate fractionated in 50 µL fractions. Fractions were analysed by SDS-PAGE with 20 µL of each fraction and 5 µL of the concentrated sample loaded; $CtIP_{FL}$ was detectable by Coomassie blue stain using the standard procedure (figure 5.2.4.3-2).

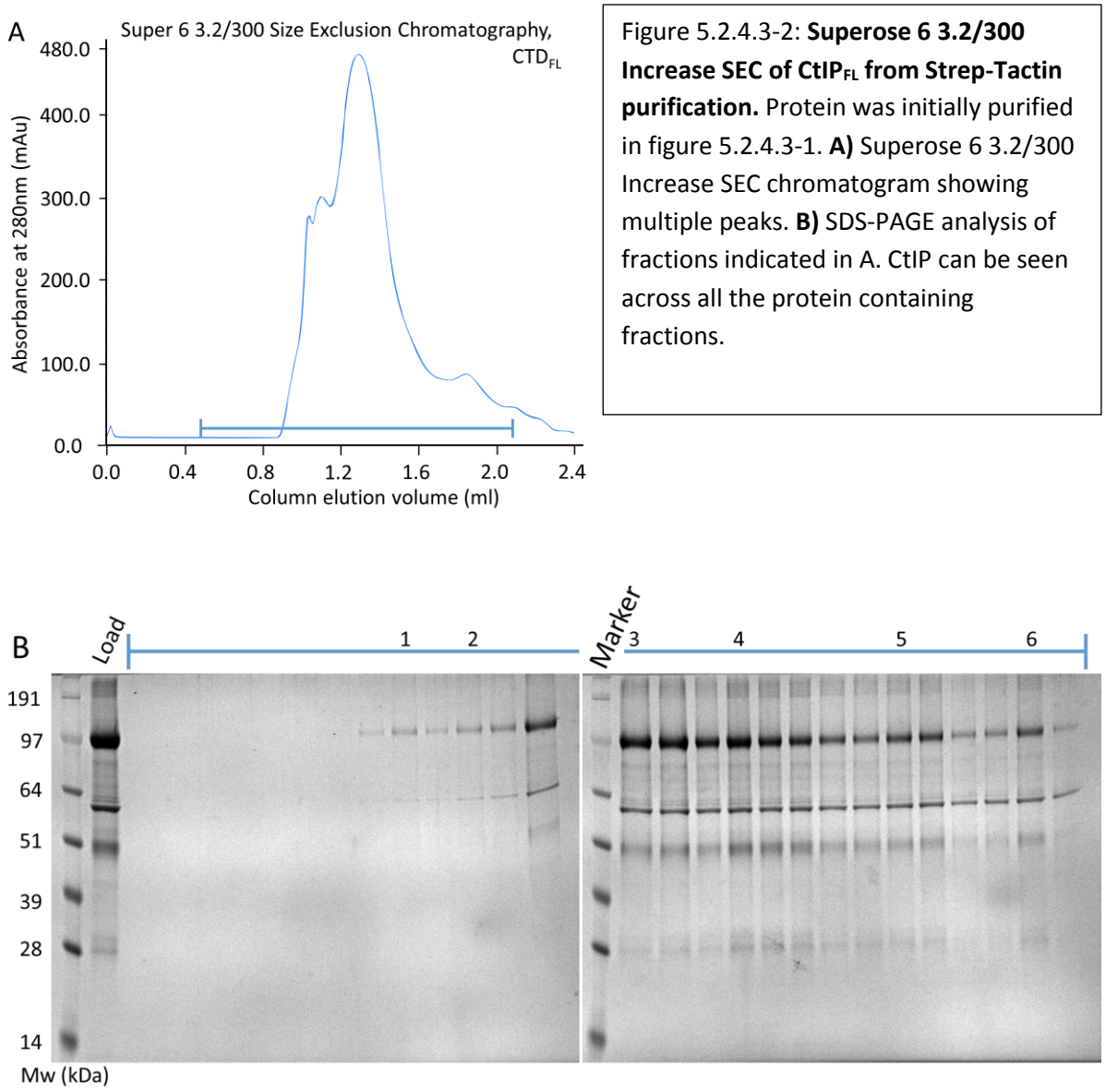


Figure 5.2.4.3-2: **Superose 6 3.2/300 Increase SEC of CtIP$_{FL}$ from Strep-Tactin purification.** Protein was initially purified in figure 5.2.4.3-1. **A)** Superose 6 3.2/300 Increase SEC chromatogram showing multiple peaks. **B)** SDS-PAGE analysis of fractions indicated in A. CtIP can be seen across all the protein containing fractions.

CtIP$_{FL}$ can be detected across all the fractions with multiple CtIP containing peaks present indicated by numbers (figure 5.2.4.3-2B). No separation of individual degradation products can be seen. Each of the peaks 3-6 contains CtIP$_{FL}$ and degraded protein, the co-migration of these degradation products shows they are forming a complex with CtIP$_{FL}$. Peaks 1&2 however mainly contain CtIP$_{FL}$ and a sub-stoichiometric level of Hsp70 protein. The elution point of the first two peaks indicates they are forming larger mega-Dalton complexes rather than the tetrameric 425 kDa CtIP$_{FL}$ complex. Although column pressure did not change over the run, washing the column with sodium hydroxide at the end of the experiment eluted further protein from the column. As the column was washed before the run, this is likely to be CtIP that could not pass through the column or bound to the Superose 6 increase matrix.

5.2.4.4 Analytical ultra-Centrifugation (AUC) analysis of $CtIP_{FL}$

Analytical Ultra Centrifugation (AUC) was performed to further analyse the different forms of $CtIP_{FL}$. This technique is one of the classical methods to characterise different aspects of proteins in solution including gross shape, weight and protein self-association (Brown et al., 2008; Cole et al., 2008).

The $CtIP_{FL}$ Twin-Strep construct was expressed in 800 mL of HEK293F cells with the soluble protein filtered and applied to a 1 mL Hi-Trap Strep-Tactin column via Superloop. The column was washed, in both forward and reverse flow before the retained Strep-tagged was $CtIP_{FL}$ eluted. Fractions were analysed by SDS-PAGE and pooled accordingly; the $CtIP_{FL}$ was then dialysed overnight into the standard buffer (figure 5.2.4.4a/b). Three samples were analysed by AUC at 100%, 50% and 25% protein concentration (figure 5.2.4.4a). The data were fitted using a continuous c(S) distribution with a frictional ratio (f/f_0) of 2.006, with various distributions of protein due to the non-homogeneous sample. Figure 5.2.4.4 shows the interferometry from the AUC experiment with panel D showing the distribution from 5S to 15S of different sedimentation values for the multimeric CtIP species. As C-terminal degradation products of CtIP can still form a tetramer there are a large array of different species within the sample, therefore a fixed molecular weight could not be calculated. Peak 1 corresponds to 45 kDa and is likely an underestimate of HSP70 present within the sample due to the high bulk frictional ratio and globular nature of HSP70. Peak 2 corresponds to between 200-240 kDa which is much larger than monomeric CtIP at 106 kDa, supporting formation of an oligomeric species with a distribution of degradation products. The sharp peak towards the end of the sedimentation coefficient graph represents larger species with sedimentation values between 20-50S. This indicates very large molecules were present in the solution such as filaments as seen in Cryo-EM (figure 5.2.4.4d).

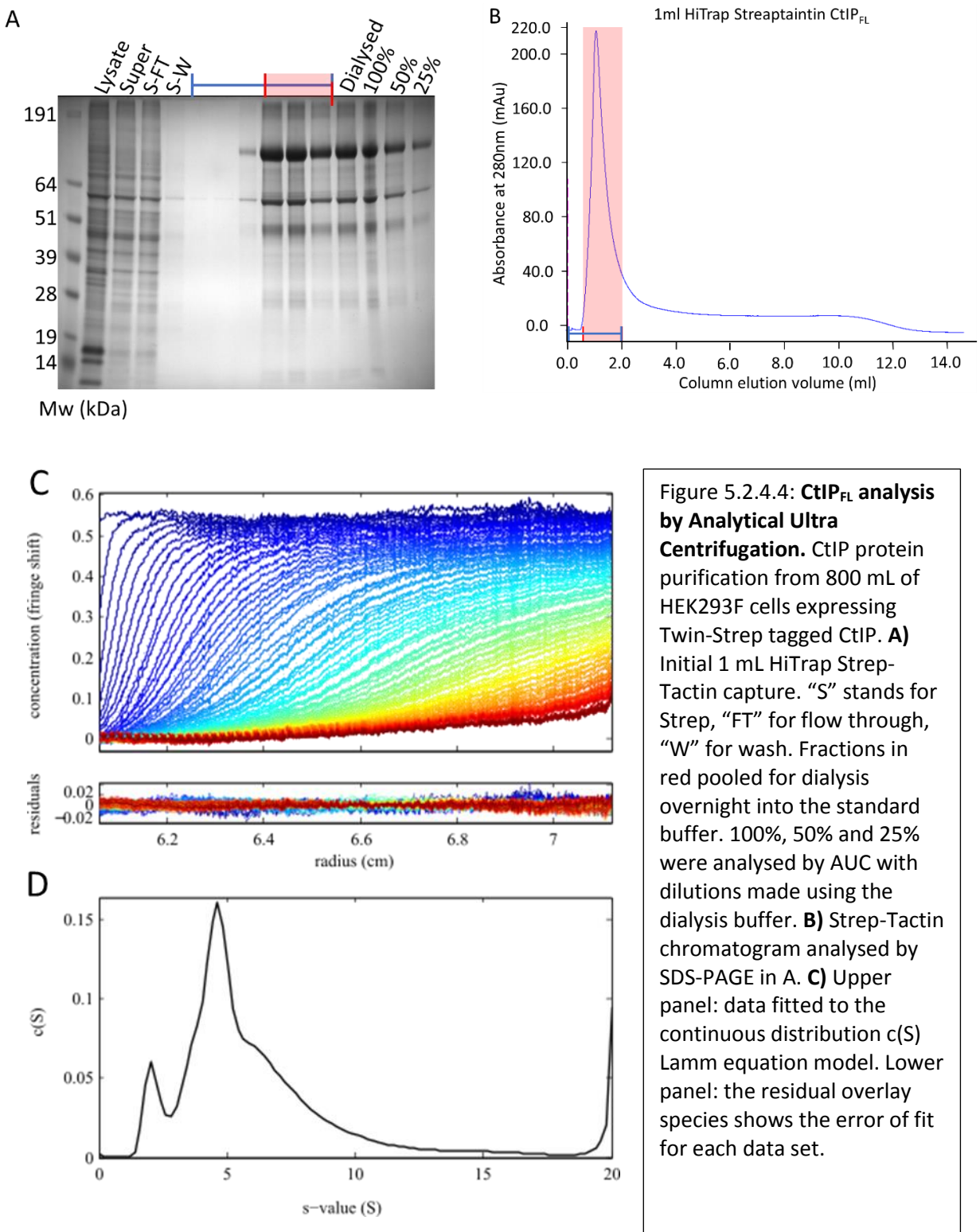


Figure 5.2.4.4: **CtIP$_{FL}$ analysis by Analytical Ultra Centrifugation.** CtIP protein purification from 800 mL of HEK293F cells expressing Twin-Strep tagged CtIP. **A)** Initial 1 mL HiTrap Strep-Tactin capture. "S" stands for Strep, "FT" for flow through, "W" for wash. Fractions in red pooled for dialysis overnight into the standard buffer. 100%, 50% and 25% were analysed by AUC with dilutions made using the dialysis buffer. **B)** Strep-Tactin chromatogram analysed by SDS-PAGE in A. **C)** Upper panel: data fitted to the continuous distribution c(S) Lamm equation model. Lower panel: the residual overlay species shows the error of fit for each data set.

D) The sedimentation coefficient distribution of all species up to 20S. Peak 1 corresponds to a weight of 45 kDa with peak 2 corresponding to between 200-240 kDa. Further species were seen at low abundance above 20S to 50S.

5.3 Conclusions and future work

In vitro biochemical studies of full length human CtIP protein have been impeded by the arduous task of purifying recombinant protein. First attempts to purify $CtIP_{FL}$ using the *E. coli* expression system gave a multitude of degradation products as shown by Western Blot analysis. $CtIP_{FL}$ expression has however been achieved in both the baculovirus and HEK293F expression systems, with $CtIP_{FL}$ protein purified to above 60% purity for biochemical analysis. Tetramerisation plays a role in the purity limitations, as C-terminal degradation products co-purify through the N-terminal oligomerisation. This has been shown by Western Blot analysis with no major difference between the two systems. Heat shock protein 70 (HSP70) has been shown to co-purify with CtIP at a sub-stoichiometric ratio as a further source of contamination. Nevertheless, the bulk purified protein has been shown to be soluble and non-aggregated with passage through the beads of both Superdex 200 and Superose 6 SEC columns. Preliminary thermal melt analysis has also determined an unfolding temperature of 42.5°C from the purified HEK293F protein. As heat shock proteins can be stable at above 50°C (Boswell-Casteel et al., 2015), this indicates the presence of stable folded $CtIP_{FL}$ domains.

Further analysis of $CtIP_{FL}$ in size determination have shown multiple species to be present. These have ranged from dimeric, tetrameric and further oligomeric species with evidence for filament formation seen in both Cryo-EM and AUC analysis. From Cryo-EM single particles are difficult to be seen, probably due to the predicted unstructured regions of CtIP, however filaments are reproducibly present within micrographs indicating a higher oligomeric state. AUC further confirms the presence of higher oligomers. The main bulk of protein (S5-S10) however gave a molecular weight between 200-240 kDa. This indicates two important features: first, this is greater than a monomer or dimeric molecular weight and therefore oligomers are formed; second, this is lower than the expected $CtIP_{FL}$ tetrameric weight indicating that intact $CtIP_{FL}$ is more likely to be within the filament at higher sedimentation values. This is further supported by Superose 6 SEC with $CtIP_{FL}$ eluting earlier with minimal co-elution of degradation products, with composite oligomers containing $CtIP_{FL}$ eluting at various volumes. A high frictional coefficient was also determined through AUC, which is higher than that of the N-terminal domain alone at 1.8 (Sun, 2013), providing further evidence that CtIP has an elongated nature.

Biochemical studies have revealed $CtIP_{FL}$ and the tetrameric mutant L27E $CtIP_{FL}$ can bind dsDNA, with a greater affinity than the CtIP C-terminal domain alone. Determination of a precise DNA binding affinity has yet to be achievable due to the inability to determine the absolute $CtIP_{FL}$ concentration. Future optimisations will be necessary for this to be achievable. Further DNA binding experiments with $CtIP_{FL}$ have shown affinity to be irrespective of breaks in the phosphodiester back bone

juxtaposed to the 5’ or 3’ strand end or the presence of ssDNA tracks on either strand mimicking resected DNA. This could be accounted for by the length of the total DNA as there is still sufficient dsDNA for a CtIP interaction. Future work within this project would involve shorter DNA constructs to determine any effect of the ssDNA tails or gaps in the DNA. To confirm $CtIP_{FL}$-DNA complex formation fluorescence polarisation (FP) could be used. The quantity of $CtIP_{FL}$ protein necessary for such experiments however is the limiting factor. Preliminary studies using the technique have shown CtIP to interact with the 96 well plate and therefore will require optimisation.

Alternative methods of purification optimisation, chromatography and experimental technique will be necessary for absolute determination of DNA binding affinity and the different oligomeric states present of $CtIP_{FL}$. These optimisations will also enable selective Cryo-EM of the different species. Future work within this area will be to apply the different fractions of $CtIP_{FL}$ from Superose 6 SEC and analyse any changes in protein seen across the elution profile. With the aim of this project area to determine a structure of $CtIP_{FL}$ by Cryo-EM. Purification of the Mre11-Rad50-Nbs1 complex has been successfully achieved with homogenous protein purified that possessed nuclease activity. This will allow for future studies between HEK293F expressed $CtIP_{FL}$ and the MRN complex. The overarching aim of this project will be to determine the structural changes brought about by different CtIP-MRN interactions through biophysical and structural methods. The starting building blocks for these future projects however have been set out within this chapter for structural analysis and conclusive oligomeric state determination.

Chapter 6: Discussion

The studies presented in this thesis characterise structural and biophysical features of the two evolutionarily conserved regions (NTD and CTD) and the full-length human CtIP protein. The NTD is known to be required for the action of DNA-end resection, with deletion studies within the domain abolishing Nbs1 binding and activation of the G2/M checkpoint. However, studies of the domain have been limited to structural determination of the tetramerisation residues 18-52 and biophysical measurements (Davies et al., 2015). Here X-ray crystal structure determination of residues 31-136 with two cysteine-to-alanine mutations has provided an insight into the domain; confirming the formation of an elongated parallel coiled coil by both crystallography and SEC-SAXS. My structure agrees with the observation that the coiled-coil domain is separated at the zinc binding site, with discontinuation of the coiled-coil into a pair of parallel alpha helices. As the NTD is known to coordinate zinc through the two cysteines, it is likely that the alanine mutations encourage the alpha helical formation in this region with a local change in conformation. Nevertheless, the coiled-coil is juxtaposed to this zinc-binding region limiting any conformational changes. SEC-SAXS measurements have been consistent between the wild type and double mutant proteins in agreement that any differences in this region are minimal. Previous models using light scattering postulated the NTD to form a four-helix bundle comprising two coiled-coils. However, the results presented here confirm the N-terminal domain forms an elongated coiled-coil. Analysis of the full-length CtIP protein by AUC further supports a long rod-shaped protein and an elongated structure.

X-ray diffraction of the wild type NTD was insufficient for structure determination, but the diffraction pattern was suggestive of a partial fibrous nature of the crystals. Fibre formation was also seen for CtIP full-length protein by Cryo-EM, with large molecular weight species on AUC and Superose 6 SEC. However, the precise mechanisms for this filament formation and the biological relevance is unknown. While CtIP Is regulated by a consortium of post translational modifications, filament formation may show a further method to regulate the protein through storage within a higher oligomeric state. This would function to protect the protein from degradation and enable a fast, non-transcriptional, response to DNA DSBs by disassembly of the filament.

Oligomerisation of CtIP at the NTD is well established for both human CtIP and the Sae2/Ctp1 orthologues. While dimerisation of the CTD has been shown necessary for interaction with LMO4, oligomerisation of the CTD remains poorly understood. Within my thesis I have consistently shown the formation of a higher molecular weight species on purification of the CTD. Chromatography techniques have been exploited to investigate the oligomeric state of the broader C-terminal region

supporting the existence of a dimeric species. While secondary structure analysis of CtIP predicted a coiled-coil motif near the C-terminus juxtaposed to the functionally-conserved homology domain. SEC-SAXS scattering analysis of this region showed a rod like structure with an elongated envelope generated, supporting the presence of a coiled-coil upstream of the CtIP Sae2 homology domain.

Both NTD and CTD possess zinc-binding regions, however the function of these two-separate zinc-binding domains is unknown. Mutation of the zinc-binding motif to alanine residues is known to abolish zinc binding within the NTD. However, the mutation showed little change to the protein structure by SEC-SAXS and enabled structure determination. Conversely, mutation of a single or pair of cysteine residues within the CTD failed to abolish zinc binding with increased degradation products when 6/8 of the purported residues were mutated. The DNA binding ability of the CTD has also been further explored by FP. Using FP, a zinc dependency has been shown, indicating that zinc could be involved in structural arrangement of this domain. Conversely, zinc is dispensable for DNA binding by EMSA and is not required for dimer formation. Both monomer and dimer of the CtIP-CTD have shown a low micromolar affinity to DNA. The DNA binding ability has been translated to that of the full-length CtIP protein with a greater affinity for DNA than the CTD, of both tetrameric CtIP and dimeric L27E CtIP. This points towards further factors within CtIP increasing DNA binding at its C-termini. Both full-length CtIP and its CTD were able to interact with the same minimal length of DNA, a value which is comparable to that obtained for the CtIP orthologue Ctp1 (Andres et al., 2015). Analysis of full-length CtIP's ability to bind DNA showed no preference for DNA with a break in the phosphodiester backbone or the presence of a single stranded DNA. The observation that both monomer and dimer of the CTD bind DNA to a similar extent indicates only one may be interacting with DNA with the other able to form protein-protein interactions such as that with And1. The different interaction region presented here between And1 and CtIP further suggests that regions outside of the C-terminus can modulate the binding within the CTD which remains to be explored.

Functionally CtIP is now well established to play a critical role in DNA DSB end resection in association with the MRN complex. The current MRN model supports two functionally distinct complex formations, one for DNA end resection and one for DNA tethering (figure 1.5.3.2-3). Substantial evidence supports the ability of the MRN complex to bridge DNA ends. A similar DNA bridging action has been reported for the CtIP orthologue Ctp1, which has been shown to bind and tether two DNA strands (Andres et al., 2015). Collectively the results presented here have revealed that CtIP has many features that would enable a similar overall architecture to MRN and Cpt1 to be achieved (figure 6). It is therefore proposed that alongside a potential CtIP DNA tethering structure, CtIP can synergise post translational modifications to change its structure resulting in different affects dependent on the modifications. This would enable the various modifications to change the interaction partners and

modulate DNA binding affinity. The findings presented within this thesis provide the foundation and the molecular building blocks required to determine the detailed mechanism of CtIP functioning in complex with MRN in DNA double strand break repair and further structural determination.

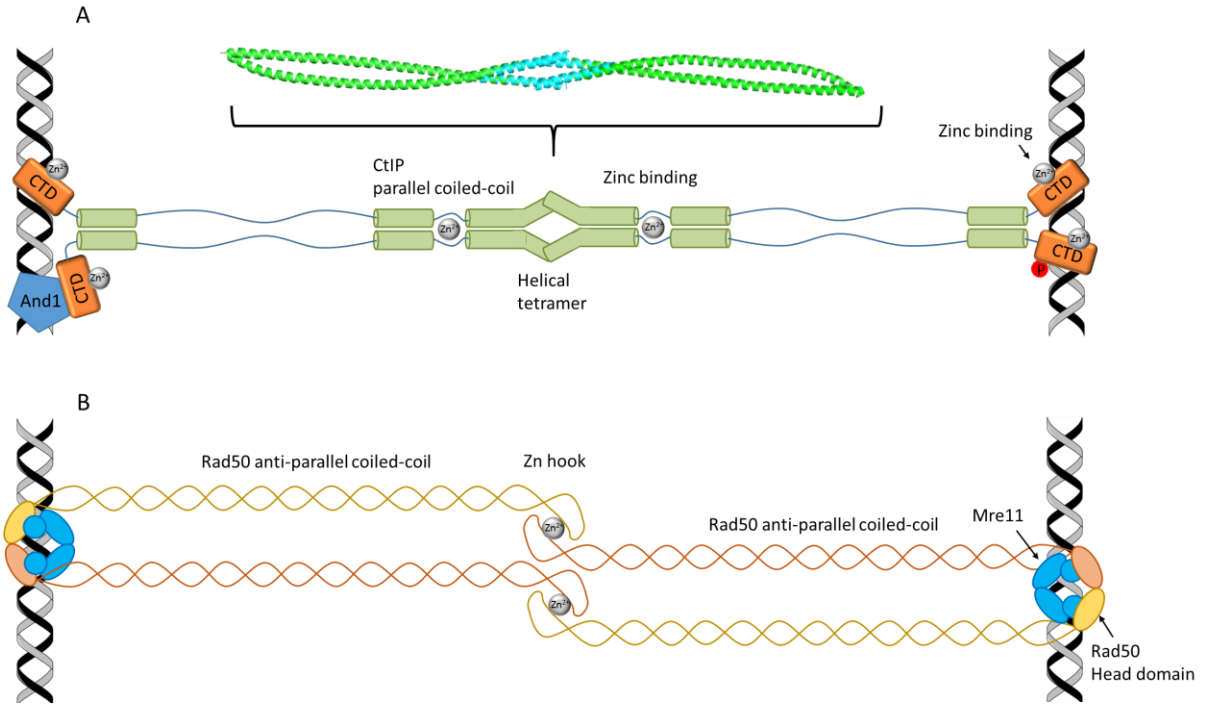


Figure 6: **Model of CtIP in comparison with the MRN complex. A)** CtIP tetramer formation is mediated by a 4 helix bundle, this extends into a parallel coiled-coil domain separated into two sections by a zinc binding site. The crystal structures of the 18-136 region are shown above the schematic representation. The middle domain of CtIP is predicted to be largely unstructured however overall the protein forms a long structure. A further coiled-coil domain precedes the conserved C-terminal region which has been shown to form a dimer, likely through this coiled-coil. The C-terminal domain can bind DNA, zinc, and contains various interaction motifs for proteins such as And1 and NBS1. This domain may function as an interaction scaffold to bring about further proteins to the MRN-CtIP complex to elicit repair. Post translational modification sites such as phosphorylation may mediate changes to DNA binding, folding or protein recruitment. **B)** The structure of Mre11-Rad50 shows the DNA tethering state which through the Rad50 coiled-coil domain and zinc hook oligomerisation allow two DNA ends to be tethered.

Appendix 1

7.1 Thermal melt buffer screen

	Thermal Shift Buffer Screen											
	1	2	3	4	5	6	7	8	9	10	11	12
A	Citrate pH 4.0 100 mM	Citrate pH 4.0 100 mM NaCl	Citrate pH 4.0 250 mM NaCl	Citrate pH 4.0 500 mM NaCl	Citrate pH 4.0 100 mM KCl	Citrate pH 4.0 250 mM KCl	Citrate pH 4.0 500 mM KCl	Citrate pH 4.0 5% DMSO 200 mM NaCl	Citrate pH 4.0 10% DMSO 200 mM NaCl	Citrate pH 4.0 10 % glycerol 200 mM NaCl	Hepes pH 7.0 100 mM MgCl2	Hepes pH 7.0 200 mM L-Arg
B	Acetate pH 5.0 100 mM	Acetate pH 5.0 100 mM NaCl	Acetate pH 5.0 250 mM NaCl	Acetate pH 5.0 500 mM NaCl	Acetate pH 5.0 100 mM KCl	Acetate pH 5.0 250 mM KCl	Acetate pH 5.0 500 mM KCl	Acetate pH 5.0 5% DMSO 200 mM NaCl	Acetate pH 5.0 10% DMSO 200 mM NaCl	Acetate pH 5.0 10 % glycerol 200 mM NaCl	Hepes pH 7.0 250 mM MgCl2	Hepes pH 7.0 400 mM L-Arg
C	MES pH 6.0 100 mM	MES pH 6.0 100 mM NaCl	MES pH 6.0 250 mM NaCl	MES pH 6.0 500 mM NaCl	MES pH 6.0 100 mM KCl	MES pH 6.0 250 mM KCl	MES pH 6.0 500 mM KCl	MES pH 6.0 5% DMSO 200 mM NaCl	MES pH 6.0 10% DMSO 200 mM NaCl	MES pH 6.0 10 % glycerol 200 mM NaCl	Hepes pH 7.0 500 mM MgCl2	Hepes pH 7.0 50 mM Arg/Glu
D	Hepes pH 7.0 100 mM	Hepes pH 7.0 100 mM NaCl	Hepes pH 7.0 250 mM NaCl	Hepes pH 7.0 500 mM NaCl	Hepes pH 7.0 100 mM KCl	Hepes pH 7.0 250 mM KCl	Hepes pH 7.0 500 mM KCl	Hepes pH 7.0 5% DMSO 200 mM NaCl	Hepes pH 7.0 10% DMSO 200 mM NaCl	Hepes pH 7.0 10 % glycerol 200 mM NaCl	Hepes pH 7.0 100 mM CaCl2	Hepes pH 7.0 100 mM NaAcetate
E	Phosphate pH 7.5 100mM	Phosphate pH 7.5 100 mM NaCl	Phosphate pH 7.5 250 mM NaCl	Phosphate pH 7.5 500 mM NaCl	Phosphate pH 7.5 100 mM KCl	Phosphate pH 7.5 250 mM KCl	Phosphate pH 7.5 500 mM KCl	Phosphate pH 7.5 5% DMSO 200 mM NaCl	Phosphate pH 7.5 10% DMSO 200 mM NaCl	Phosphate pH 7.5 10 % glycerol 200 mM NaCl	Hepes pH 7.0 100 mM Li2SO4	Hepes pH 7.0 100 mM NaFormate
F	tris pH 8.0 100 mM	tris pH 8.0 100 mM NaCl	tris pH 8.0 250 mM NaCl	tris pH 8.0 500 mM NaCl	tris pH 8.0 100 mM KCl	tris pH 8.0 250 mM KCl	tris pH 8.0 500 mM KCl	tris pH 8.0 5% DMSO 200 mM NaCl	tris pH 8.0 10% DMSO 200 mM NaCl	tris pH 8.0 10 % glycerol 200 mM NaCl	Hepes pH 7.0 100 mM (NH4)2SO4	Hepes pH 7.0 100 mM Na3Citrate
G	CHES pH 9.0 100 mM	CHES pH 9.0 100 mM NaCl	CHES pH 9.0 250 mM NaCl	CHES pH 9.0 500 mM NaCl	CHES pH 9.0 100 mM KCl	CHES pH 9.0 250 mM KCl	CHES pH 9.0 500 mM KCl	CHES pH 9.0 5% DMSO 200 mM NaCl	CHES pH 9.0 10% DMSO 200 mM NaCl	CHES pH 9.0 10 % glycerol 200 mM NaCl	Hepes pH 7.0 100 mM NH4C2H3O2	Hepes pH 7.0 5% PEG 4,000
H	CAPS pH 10.0 100 mM	CAPS pH 10.0 100 mM NaCl	CAPS pH 10.0 250 mM NaCl	CAPS pH 10.0 500 mM NaCl	CAPS pH 10.0 100 mM KCl	CAPS pH 10.0 250 mM KCl	CAPS pH 10.0 500 mM KCl	CAPS pH 10.0 5% DMSO 200 mM NaCl	CAPS pH 10.0 10% DMSO 200 mM NaCl	CAPS pH 10.0 10 % glycerol 200 mM NaCl	Hepes pH 7.0 100 mM Mg(C2H3O2)2	Hepes pH 7.0 10% PEG 4000

Supplementary figure 7.1: **Hyvönen group thermal shift buffer screen.** Buffers were mixed 50:50 to NTD protein for thermal melt analysis in figure 3.2.2.1.

7.2 Optimisation 1

Component Name	Stock Concentration	Units	Type
[1]Amonium sulphate	3000	mM	Precipitant
[2]Mes (max pH range: 5.5-6.7)			Buffer
- Mes (pH: 5.5)	500.00	mM	
- Mes (pH: 6.7)	500.00	mM	
[3]Hepes (max pH range: 6.8-8.2)			Buffer
- Hepes (pH: 6.8)	500.00	mM	
- Hepes (pH: 8.2)	500.00	mM	
[4]Tris (max pH range: 7-9)			Buffer
- Tris (pH: 7)	500.00	mM	
- Tris (pH: 9)	500.00	mM	
[5]CHES (max pH range: 8.6-10)			Buffer
- CHES (pH: 8.6)	500.00	mM	
- CHES (pH: 10)	500.00	mM	

Well volume 85 µL

	1	2	3	4	5	6	7	8	9	10	11	12
A	[1]: 400.00mM [2]: 100.00uL, pH 6.00	[1]: 500.00mM [2]: 100.00uL, pH 6.00	[1]: 600.00mM [2]: 100.00uL, pH 6.00	[1]: 700.00mM [2]: 100.00uL, pH 6.00	[1]: 800.00mM [2]: 100.00uL, pH 6.00	[1]: 900.00mM [2]: 100.00uL, pH 6.00	[1]: 1000.00mM [2]: 100.00uL, pH 6.00	[1]: 1100.00mM [2]: 100.00uL, pH 6.00	[1]: 1200.00mM [2]: 100.00uL, pH 6.00	[1]: 1300.00mM [2]: 100.00uL, pH 6.00	[1]: 1400.00mM [2]: 100.00uL, pH 6.00	[1]: 1500.00mM [2]: 100.00uL, pH 6.00
B	[1]: 400.00mM [2]: 100.00uL, pH 6.50	[1]: 500.00mM [2]: 100.00uL, pH 6.50	[1]: 600.00mM [2]: 100.00uL, pH 6.50	[1]: 700.00mM [2]: 100.00uL, pH 6.50	[1]: 800.00mM [2]: 100.00uL, pH 6.50	[1]: 900.00mM [2]: 100.00uL, pH 6.50	[1]: 1000.00mM [2]: 100.00uL, pH 6.50	[1]: 1100.00mM [2]: 100.00uL, pH 6.50	[1]: 1200.00mM [2]: 100.00uL, pH 6.50	[1]: 1300.00mM [2]: 100.00uL, pH 6.50	[1]: 1400.00mM [2]: 100.00uL, pH 6.50	[1]: 1500.00mM [2]: 100.00uL, pH 6.50
C	[1]: 400.00mM [3]: 100.00uL, pH 7.00	[1]: 500.00mM [3]: 100.00uL, pH 7.00	[1]: 600.00mM [3]: 100.00uL, pH 7.00	[1]: 700.00mM [3]: 100.00uL, pH 7.00	[1]: 800.00mM [3]: 100.00uL, pH 7.00	[1]: 900.00mM [3]: 100.00uL, pH 7.00	[1]: 1000.00mM [3]: 100.00uL, pH 7.00	[1]: 1100.00mM [3]: 100.00uL, pH 7.00	[1]: 1200.00mM [3]: 100.00uL, pH 7.00	[1]: 1300.00mM [3]: 100.00uL, pH 7.00	[1]: 1400.00mM [3]: 100.00uL, pH 7.00	[1]: 1500.00mM [3]: 100.00uL, pH 7.00
D	[1]: 400.00mM [3]: 100.00uL, pH 7.50	[1]: 500.00mM [3]: 100.00uL, pH 7.50	[1]: 600.00mM [3]: 100.00uL, pH 7.50	[1]: 700.00mM [3]: 100.00uL, pH 7.50	[1]: 800.00mM [3]: 100.00uL, pH 7.50	[1]: 900.00mM [3]: 100.00uL, pH 7.50	[1]: 1000.00mM [3]: 100.00uL, pH 7.50	[1]: 1100.00mM [3]: 100.00uL, pH 7.50	[1]: 1200.00mM [3]: 100.00uL, pH 7.50	[1]: 1300.00mM [3]: 100.00uL, pH 7.50	[1]: 1400.00mM [3]: 100.00uL, pH 7.50	[1]: 1500.00mM [3]: 100.00uL, pH 7.50
E	[1]: 400.00mM [4]: 100.00uL, pH 8.00	[1]: 500.00mM [4]: 100.00uL, pH 8.00	[1]: 600.00mM [4]: 100.00uL, pH 8.00	[1]: 700.00mM [4]: 100.00uL, pH 8.00	[1]: 800.00mM [4]: 100.00uL, pH 8.00	[1]: 900.00mM [4]: 100.00uL, pH 8.00	[1]: 1000.00mM [4]: 100.00uL, pH 8.00	[1]: 1100.00mM [4]: 100.00uL, pH 8.00	[1]: 1200.00mM [4]: 100.00uL, pH 8.00	[1]: 1300.00mM [4]: 100.00uL, pH 8.00	[1]: 1400.00mM [4]: 100.00uL, pH 8.00	[1]: 1500.00mM [4]: 100.00uL, pH 8.00
F	[1]: 400.00mM [4]: 100.00uL, pH 8.50	[1]: 500.00mM [4]: 100.00uL, pH 8.50	[1]: 600.00mM [4]: 100.00uL, pH 8.50	[1]: 700.00mM [4]: 100.00uL, pH 8.50	[1]: 800.00mM [4]: 100.00uL, pH 8.50	[1]: 900.00mM [4]: 100.00uL, pH 8.50	[1]: 1000.00mM [4]: 100.00uL, pH 8.50	[1]: 1100.00mM [4]: 100.00uL, pH 8.50	[1]: 1200.00mM [4]: 100.00uL, pH 8.50	[1]: 1300.00mM [4]: 100.00uL, pH 8.50	[1]: 1400.00mM [4]: 100.00uL, pH 8.50	[1]: 1500.00mM [4]: 100.00uL, pH 8.50
G	[1]: 400.00mM [5]: 100.00uL, pH 9.00	[1]: 500.00mM [5]: 100.00uL, pH 9.00	[1]: 600.00mM [5]: 100.00uL, pH 9.00	[1]: 700.00mM [5]: 100.00uL, pH 9.00	[1]: 800.00mM [5]: 100.00uL, pH 9.00	[1]: 900.00mM [5]: 100.00uL, pH 9.00	[1]: 1000.00mM [5]: 100.00uL, pH 9.00	[1]: 1100.00mM [5]: 100.00uL, pH 9.00	[1]: 1200.00mM [5]: 100.00uL, pH 9.00	[1]: 1300.00mM [5]: 100.00uL, pH 9.00	[1]: 1400.00mM [5]: 100.00uL, pH 9.00	[1]: 1500.00mM [5]: 100.00uL, pH 9.00
H	[1]: 400.00mM [5]: 100.00uL, pH 9.50	[1]: 500.00mM [5]: 100.00uL, pH 9.50	[1]: 600.00mM [5]: 100.00uL, pH 9.50	[1]: 700.00mM [5]: 100.00uL, pH 9.50	[1]: 800.00mM [5]: 100.00uL, pH 9.50	[1]: 900.00mM [5]: 100.00uL, pH 9.50	[1]: 1000.00mM [5]: 100.00uL, pH 9.50	[1]: 1100.00mM [5]: 100.00uL, pH 9.50	[1]: 1200.00mM [5]: 100.00uL, pH 9.50	[1]: 1300.00mM [5]: 100.00uL, pH 9.50	[1]: 1400.00mM [5]: 100.00uL, pH 9.50	[1]: 1500.00mM [5]: 100.00uL, pH 9.50

Supplementary figure 7.2: **Optimisation 1 - 96 well condition screen.**

7.3 Optimisation 18 NaSCN

Component Name	Stock Concentration	Units	Type
[1]Amonium sulphate	3000	mM	Precipitant
[2]glycerol	50	%v/v	[Unspecified]
[3]Mes (max pH range: 5.5-6.7) - Mes (pH: 5.5) - Mes (pH: 6.7)	 500.00 500.00	 mM mM	Buffer

Well volume 85 µL

	1	2	3	4	5	6	7	8	9	10	11	12
A	[1]: 1425.00mM [2]: 6.50%v/v [3]: 100.00uL pH 5.90	[1]: 1425.00mM [2]: 6.50%v/v [3]: 100.00uL pH 5.90	[1]: 1425.00mM [2]: 6.50%v/v [3]: 100.00uL pH 5.90	[1]: 1425.00mM [2]: 6.50%v/v [3]: 100.00uL pH 5.90	[1]: 1425.00mM [2]: 6.50%v/v [3]: 100.00uL pH 5.90	[1]: 1425.00mM [2]: 6.50%v/v [3]: 100.00uL pH 5.90	[1]: 1425.00mM [2]: 6.50%v/v [3]: 100.00uL pH 5.90	[1]: 1425.00mM [2]: 6.50%v/v [3]: 100.00uL pH 5.90	[1]: 1425.00mM [2]: 6.50%v/v [3]: 100.00uL pH 5.90	[1]: 1425.00mM [2]: 6.50%v/v [3]: 100.00uL pH 5.90	[1]: 1425.00mM [2]: 6.50%v/v [3]: 100.00uL pH 5.90	[1]: 1425.00mM [2]: 6.50%v/v [3]: 100.00uL pH 5.90
B	[1]: 1450.00mM [2]: 6.50%v/v [3]: 100.00uL pH 5.90	[1]: 1450.00mM [2]: 6.50%v/v [3]: 100.00uL pH 5.90	[1]: 1450.00mM [2]: 6.50%v/v [3]: 100.00uL pH 5.90	[1]: 1450.00mM [2]: 6.50%v/v [3]: 100.00uL pH 5.90	[1]: 1450.00mM [2]: 6.50%v/v [3]: 100.00uL pH 5.90	[1]: 1450.00mM [2]: 6.50%v/v [3]: 100.00uL pH 5.90	[1]: 1450.00mM [2]: 6.50%v/v [3]: 100.00uL pH 5.90	[1]: 1450.00mM [2]: 6.50%v/v [3]: 100.00uL pH 5.90	[1]: 1450.00mM [2]: 6.50%v/v [3]: 100.00uL pH 5.90	[1]: 1450.00mM [2]: 6.50%v/v [3]: 100.00uL pH 5.90	[1]: 1450.00mM [2]: 6.50%v/v [3]: 100.00uL pH 5.90	[1]: 1450.00mM [2]: 6.50%v/v [3]: 100.00uL pH 5.90
C	[1]: 1475.00mM [2]: 6.50%v/v [3]: 100.00uL pH 5.90	[1]: 1475.00mM [2]: 6.50%v/v [3]: 100.00uL pH 5.90	[1]: 1475.00mM [2]: 6.50%v/v [3]: 100.00uL pH 5.90	[1]: 1475.00mM [2]: 6.50%v/v [3]: 100.00uL pH 5.90	[1]: 1475.00mM [2]: 6.50%v/v [3]: 100.00uL pH 5.90	[1]: 1475.00mM [2]: 6.50%v/v [3]: 100.00uL pH 5.90	[1]: 1475.00mM [2]: 6.50%v/v [3]: 100.00uL pH 5.90	[1]: 1475.00mM [2]: 6.50%v/v [3]: 100.00uL pH 5.90	[1]: 1475.00mM [2]: 6.50%v/v [3]: 100.00uL pH 5.90	[1]: 1475.00mM [2]: 6.50%v/v [3]: 100.00uL pH 5.90	[1]: 1475.00mM [2]: 6.50%v/v [3]: 100.00uL pH 5.90	[1]: 1475.00mM [2]: 6.50%v/v [3]: 100.00uL pH 5.90
D	[1]: 1500.00mM [2]: 6.50%v/v [3]: 100.00uL pH 5.90	[1]: 1500.00mM [2]: 6.50%v/v [3]: 100.00uL pH 5.90	[1]: 1500.00mM [2]: 6.50%v/v [3]: 100.00uL pH 5.90	[1]: 1500.00mM [2]: 6.50%v/v [3]: 100.00uL pH 5.90	[1]: 1500.00mM [2]: 6.50%v/v [3]: 100.00uL pH 5.90	[1]: 1500.00mM [2]: 6.50%v/v [3]: 100.00uL pH 5.90	[1]: 1500.00mM [2]: 6.50%v/v [3]: 100.00uL pH 5.90	[1]: 1500.00mM [2]: 6.50%v/v [3]: 100.00uL pH 5.90	[1]: 1500.00mM [2]: 6.50%v/v [3]: 100.00uL pH 5.90	[1]: 1500.00mM [2]: 6.50%v/v [3]: 100.00uL pH 5.90	[1]: 1500.00mM [2]: 6.50%v/v [3]: 100.00uL pH 5.90	[1]: 1500.00mM [2]: 6.50%v/v [3]: 100.00uL pH 5.90
E	[1]: 1525.00mM [2]: 6.50%v/v [3]: 100.00uL pH 5.90	[1]: 1525.00mM [2]: 6.50%v/v [3]: 100.00uL pH 5.90	[1]: 1525.00mM [2]: 6.50%v/v [3]: 100.00uL pH 5.90	[1]: 1525.00mM [2]: 6.50%v/v [3]: 100.00uL pH 5.90	[1]: 1525.00mM [2]: 6.50%v/v [3]: 100.00uL pH 5.90	[1]: 1525.00mM [2]: 6.50%v/v [3]: 100.00uL pH 5.90	[1]: 1525.00mM [2]: 6.50%v/v [3]: 100.00uL pH 5.90	[1]: 1525.00mM [2]: 6.50%v/v [3]: 100.00uL pH 5.90	[1]: 1525.00mM [2]: 6.50%v/v [3]: 100.00uL pH 5.90	[1]: 1525.00mM [2]: 6.50%v/v [3]: 100.00uL pH 5.90	[1]: 1525.00mM [2]: 6.50%v/v [3]: 100.00uL pH 5.90	[1]: 1525.00mM [2]: 6.50%v/v [3]: 100.00uL pH 5.90
F	[1]: 1550.00mM [2]: 6.50%v/v [3]: 100.00uL pH 5.90	[1]: 1550.00mM [2]: 6.50%v/v [3]: 100.00uL pH 5.90	[1]: 1550.00mM [2]: 6.50%v/v [3]: 100.00uL pH 5.90	[1]: 1550.00mM [2]: 6.50%v/v [3]: 100.00uL pH 5.90	[1]: 1550.00mM [2]: 6.50%v/v [3]: 100.00uL pH 5.90	[1]: 1550.00mM [2]: 6.50%v/v [3]: 100.00uL pH 5.90	[1]: 1550.00mM [2]: 6.50%v/v [3]: 100.00uL pH 5.90	[1]: 1550.00mM [2]: 6.50%v/v [3]: 100.00uL pH 5.90	[1]: 1550.00mM [2]: 6.50%v/v [3]: 100.00uL pH 5.90	[1]: 1550.00mM [2]: 6.50%v/v [3]: 100.00uL pH 5.90	[1]: 1550.00mM [2]: 6.50%v/v [3]: 100.00uL pH 5.90	[1]: 1550.00mM [2]: 6.50%v/v [3]: 100.00uL pH 5.90
G	[1]: 1575.00mM [2]: 6.50%v/v [3]: 100.00uL pH 5.90	[1]: 1575.00mM [2]: 6.50%v/v [3]: 100.00uL pH 5.90	[1]: 1575.00mM [2]: 6.50%v/v [3]: 100.00uL pH 5.90	[1]: 1575.00mM [2]: 6.50%v/v [3]: 100.00uL pH 5.90	[1]: 1575.00mM [2]: 6.50%v/v [3]: 100.00uL pH 5.90	[1]: 1575.00mM [2]: 6.50%v/v [3]: 100.00uL pH 5.90	[1]: 1575.00mM [2]: 6.50%v/v [3]: 100.00uL pH 5.90	[1]: 1575.00mM [2]: 6.50%v/v [3]: 100.00uL pH 5.90	[1]: 1575.00mM [2]: 6.50%v/v [3]: 100.00uL pH 5.90	[1]: 1575.00mM [2]: 6.50%v/v [3]: 100.00uL pH 5.90	[1]: 1575.00mM [2]: 6.50%v/v [3]: 100.00uL pH 5.90	[1]: 1575.00mM [2]: 6.50%v/v [3]: 100.00uL pH 5.90
H	[1]: 1600.00mM [2]: 6.50%v/v [3]: 100.00uL pH 5.90	[1]: 1600.00mM [2]: 6.50%v/v [3]: 100.00uL pH 5.90	[1]: 1600.00mM [2]: 6.50%v/v [3]: 100.00uL pH 5.90	[1]: 1600.00mM [2]: 6.50%v/v [3]: 100.00uL pH 5.90	[1]: 1600.00mM [2]: 6.50%v/v [3]: 100.00uL pH 5.90	[1]: 1600.00mM [2]: 6.50%v/v [3]: 100.00uL pH 5.90	[1]: 1600.00mM [2]: 6.50%v/v [3]: 100.00uL pH 5.90	[1]: 1600.00mM [2]: 6.50%v/v [3]: 100.00uL pH 5.90	[1]: 1600.00mM [2]: 6.50%v/v [3]: 100.00uL pH 5.90	[1]: 1600.00mM [2]: 6.50%v/v [3]: 100.00uL pH 5.90	[1]: 1600.00mM [2]: 6.50%v/v [3]: 100.00uL pH 5.90	[1]: 1600.00mM [2]: 6.50%v/v [3]: 100.00uL pH 5.90

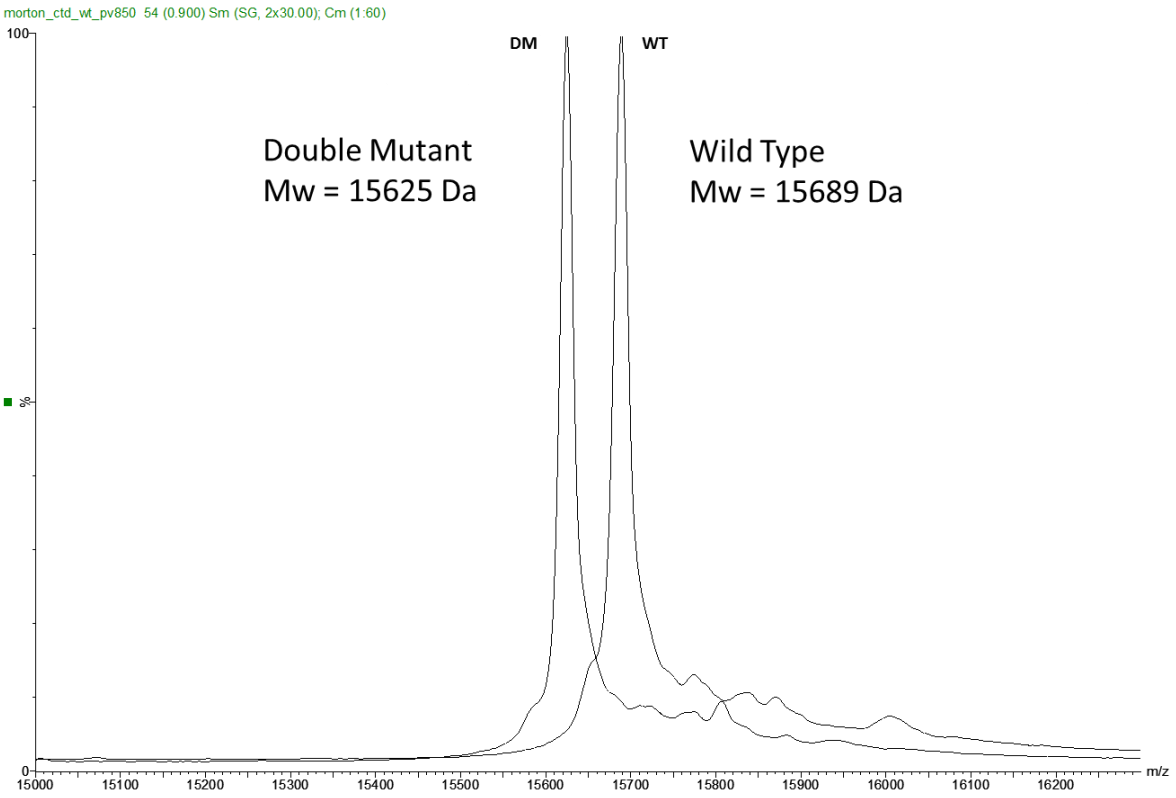
Component Name	Stock Concentration	Units	Type
[1]sodium thiocyanate	2000	mM	[Unspecified]

Well Volume: 8.5uL

	1	2	3	4	5	6	7	8	9	10	11	12
A		Diluent: 8.50uL	[1]: 0.85uL, 200.00mM Diluent: 7.65uL	[1]: 1.70uL, 400.00mM Diluent: 6.80uL	[1]: 2.55uL, 600.00mM Diluent: 5.95uL	[1]: 3.40uL, 800.00mM Diluent: 5.10uL	[1]: 4.25uL, 1000.00mM Diluent: 4.25uL	[1]: 5.10uL, 1200.00mM Diluent: 3.40uL	[1]: 5.95uL, 1400.00mM Diluent: 2.55uL	[1]: 6.80uL, 1600.00mM Diluent: 1.70uL	[1]: 7.65uL, 1800.00mM Diluent: 0.85uL	[1]: 8.50uL, 2000.00mM
B		Diluent: 8.50uL	[1]: 0.85uL, 200.00mM Diluent: 7.65uL	[1]: 1.70uL, 400.00mM Diluent: 6.80uL	[1]: 2.55uL, 600.00mM Diluent: 5.95uL	[1]: 3.40uL, 800.00mM Diluent: 5.10uL	[1]: 4.25uL, 1000.00mM Diluent: 4.25uL	[1]: 5.10uL, 1200.00mM Diluent: 3.40uL	[1]: 5.95uL, 1400.00mM Diluent: 2.55uL	[1]: 6.80uL, 1600.00mM Diluent: 1.70uL	[1]: 7.65uL, 1800.00mM Diluent: 0.85uL	[1]: 8.50uL, 2000.00mM
C		Diluent: 8.50uL	[1]: 0.85uL, 200.00mM Diluent: 7.65uL	[1]: 1.70uL, 400.00mM Diluent: 6.80uL	[1]: 2.55uL, 600.00mM Diluent: 5.95uL	[1]: 3.40uL, 800.00mM Diluent: 5.10uL	[1]: 4.25uL, 1000.00mM Diluent: 4.25uL	[1]: 5.10uL, 1200.00mM Diluent: 3.40uL	[1]: 5.95uL, 1400.00mM Diluent: 2.55uL	[1]: 6.80uL, 1600.00mM Diluent: 1.70uL	[1]: 7.65uL, 1800.00mM Diluent: 0.85uL	[1]: 8.50uL, 2000.00mM
D		Diluent: 8.50uL	[1]: 0.85uL, 200.00mM Diluent: 7.65uL	[1]: 1.70uL, 400.00mM Diluent: 6.80uL	[1]: 2.55uL, 600.00mM Diluent: 5.95uL	[1]: 3.40uL, 800.00mM Diluent: 5.10uL	[1]: 4.25uL, 1000.00mM Diluent: 4.25uL	[1]: 5.10uL, 1200.00mM Diluent: 3.40uL	[1]: 5.95uL, 1400.00mM Diluent: 2.55uL	[1]: 6.80uL, 1600.00mM Diluent: 1.70uL	[1]: 7.65uL, 1800.00mM Diluent: 0.85uL	[1]: 8.50uL, 2000.00mM
E		Diluent: 8.50uL	[1]: 0.85uL, 200.00mM Diluent: 7.65uL	[1]: 1.70uL, 400.00mM Diluent: 6.80uL	[1]: 2.55uL, 600.00mM Diluent: 5.95uL	[1]: 3.40uL, 800.00mM Diluent: 5.10uL	[1]: 4.25uL, 1000.00mM Diluent: 4.25uL	[1]: 5.10uL, 1200.00mM Diluent: 3.40uL	[1]: 5.95uL, 1400.00mM Diluent: 2.55uL	[1]: 6.80uL, 1600.00mM Diluent: 1.70uL	[1]: 7.65uL, 1800.00mM Diluent: 0.85uL	[1]: 8.50uL, 2000.00mM
F		Diluent: 8.50uL	[1]: 0.85uL, 200.00mM Diluent: 7.65uL	[1]: 1.70uL, 400.00mM Diluent: 6.80uL	[1]: 2.55uL, 600.00mM Diluent: 5.95uL	[1]: 3.40uL, 800.00mM Diluent: 5.10uL	[1]: 4.25uL, 1000.00mM Diluent: 4.25uL	[1]: 5.10uL, 1200.00mM Diluent: 3.40uL	[1]: 5.95uL, 1400.00mM Diluent: 2.55uL	[1]: 6.80uL, 1600.00mM Diluent: 1.70uL	[1]: 7.65uL, 1800.00mM Diluent: 0.85uL	[1]: 8.50uL, 2000.00mM
G		Diluent: 8.50uL	[1]: 0.85uL, 200.00mM Diluent: 7.65uL	[1]: 1.70uL, 400.00mM Diluent: 6.80uL	[1]: 2.55uL, 600.00mM Diluent: 5.95uL	[1]: 3.40uL, 800.00mM Diluent: 5.10uL	[1]: 4.25uL, 1000.00mM Diluent: 4.25uL	[1]: 5.10uL, 1200.00mM Diluent: 3.40uL	[1]: 5.95uL, 1400.00mM Diluent: 2.55uL	[1]: 6.80uL, 1600.00mM Diluent: 1.70uL	[1]: 7.65uL, 1800.00mM Diluent: 0.85uL	[1]: 8.50uL, 2000.00mM
H		Diluent: 8.50uL	[1]: 0.85uL, 200.00mM Diluent: 7.65uL	[1]: 1.70uL, 400.00mM Diluent: 6.80uL	[1]: 2.55uL, 600.00mM Diluent: 5.95uL	[1]: 3.40uL, 800.00mM Diluent: 5.10uL	[1]: 4.25uL, 1000.00mM Diluent: 4.25uL	[1]: 5.10uL, 1200.00mM Diluent: 3.40uL	[1]: 5.95uL, 1400.00mM Diluent: 2.55uL	[1]: 6.80uL, 1600.00mM Diluent: 1.70uL	[1]: 7.65uL, 1800.00mM Diluent: 0.85uL	[1]: 8.50uL, 2000.00mM

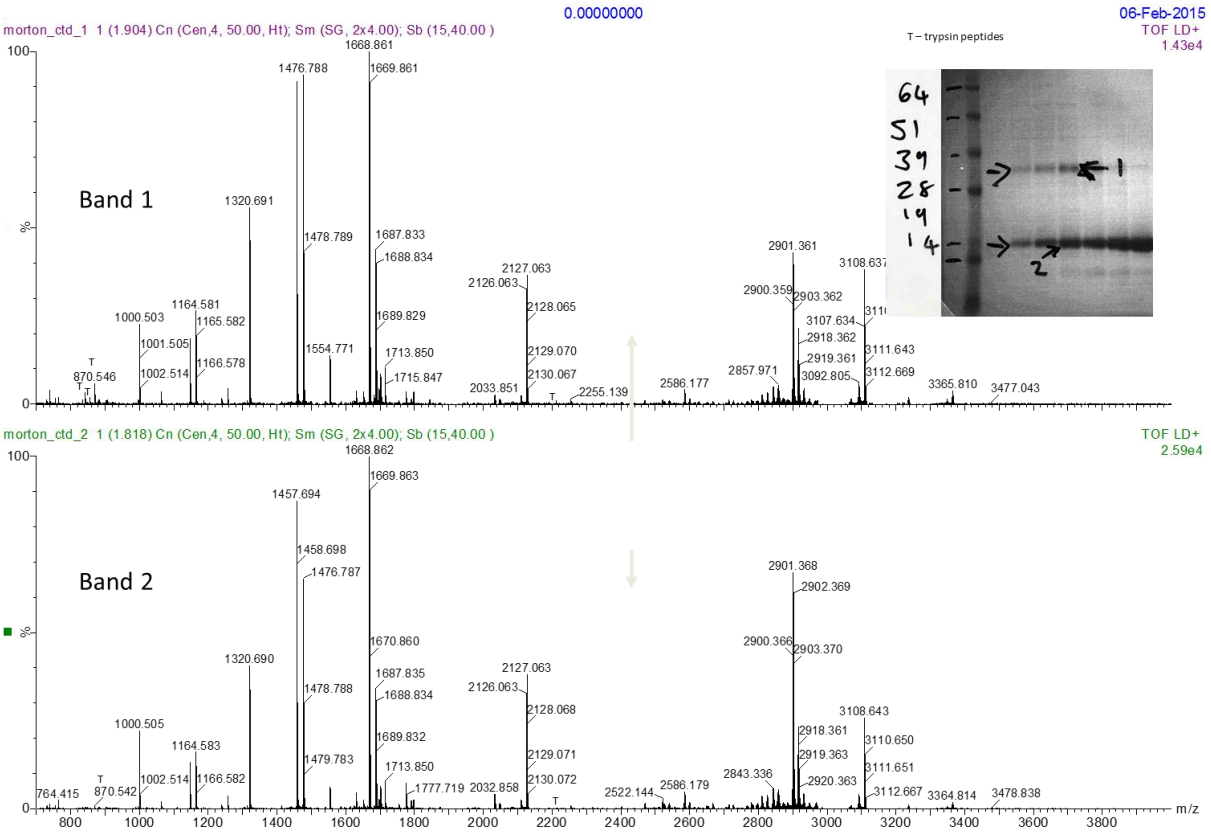
Supplementary figure 7.3: **Optimisation 18 - 96 well condition screen.**

7.4 Incorporation of double alanine mutation in CTD_{DM}

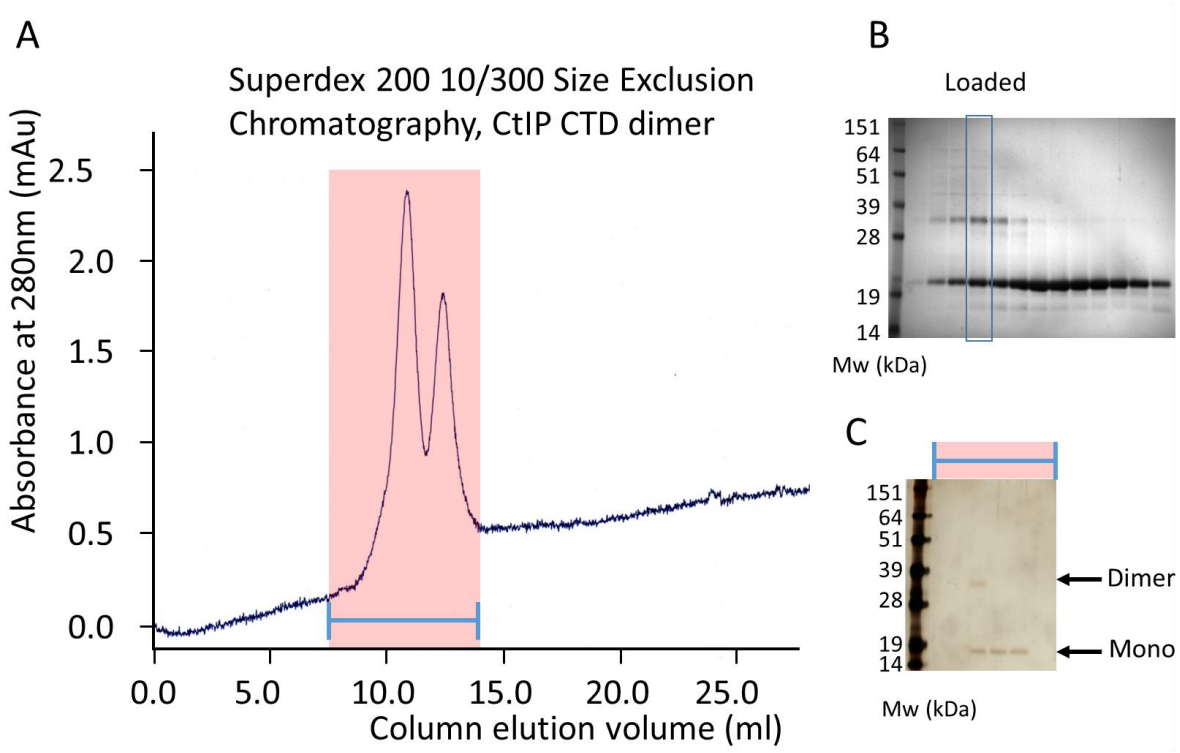


Supplementary figure 7.4: **MALDI analysis intact mass of CtIP-CTD vs CTD_{DM}**. Confirmation that the CTD_{DM} contains two alanine substitutions.

7.5 CtIP CTD Dimer MALDI-TOF – supplementary figures



Supplementary figure 7.5-1: **MALDI-TOF confirmation that CtIP-CTD can exist as a dimer in SDS-PAGE.** The same profile means that it is the same protein, consistent with band 1 being a dimer of band 2.



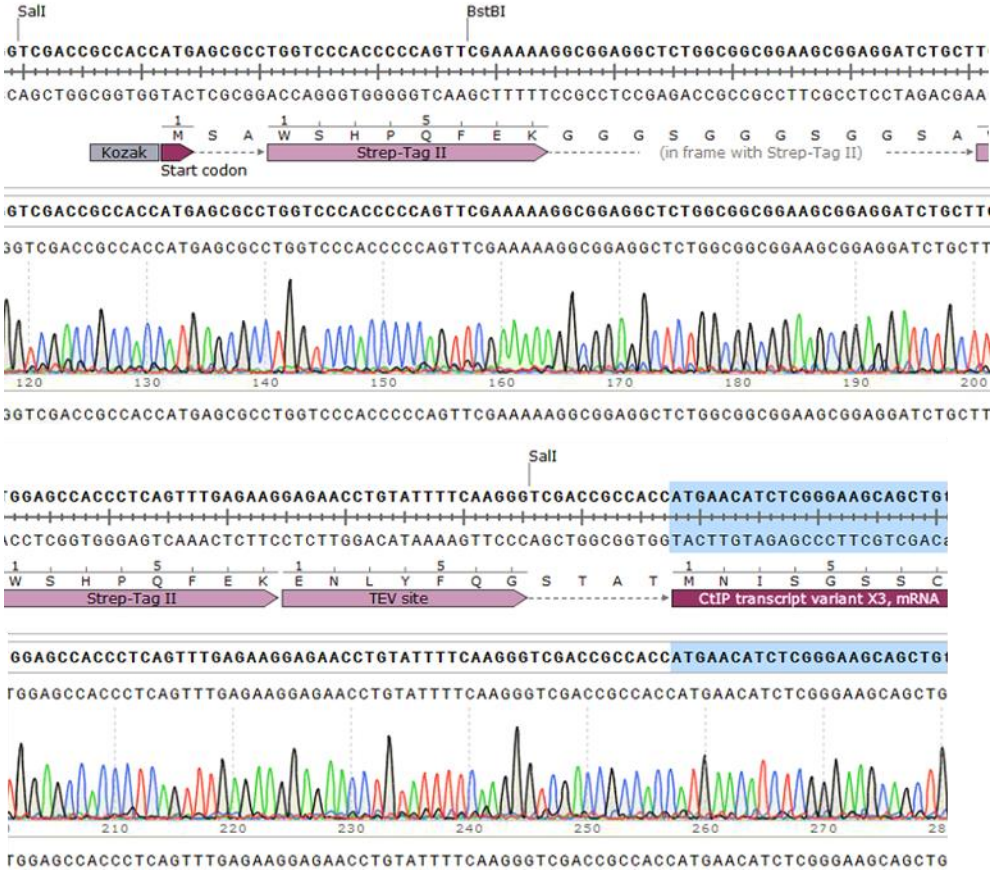
Supplementary figure 7.5-2: **SEC on dimer fraction shows both monomer and dimer peaks.** Superdex 200 10/300 SEC on CTD fraction confirmed to contain monomer and dimer. Two peaks correspond to weights for dimer and monomer (mono) respectively. **A)** SEC trace, fractions highlighted were analysed by SDS-PAGE. **B)** Fraction loaded onto SEC column. **C)** Silver stain gel of fractions in A, Dimer and Monomer bands are identified.

7.6 Antibodies

Antibody	Manufacturer	Catalogue number	Species
Anti-Myc tag	Sigma	06-549	Rabbit
Anti-His tag	Sigma	H1029	mouse
Anti-Flag M2 monoclonal	Sigma	F1804	Mouse
Anti-Strep tag II	RayBiotech	168-10750	Rabbit

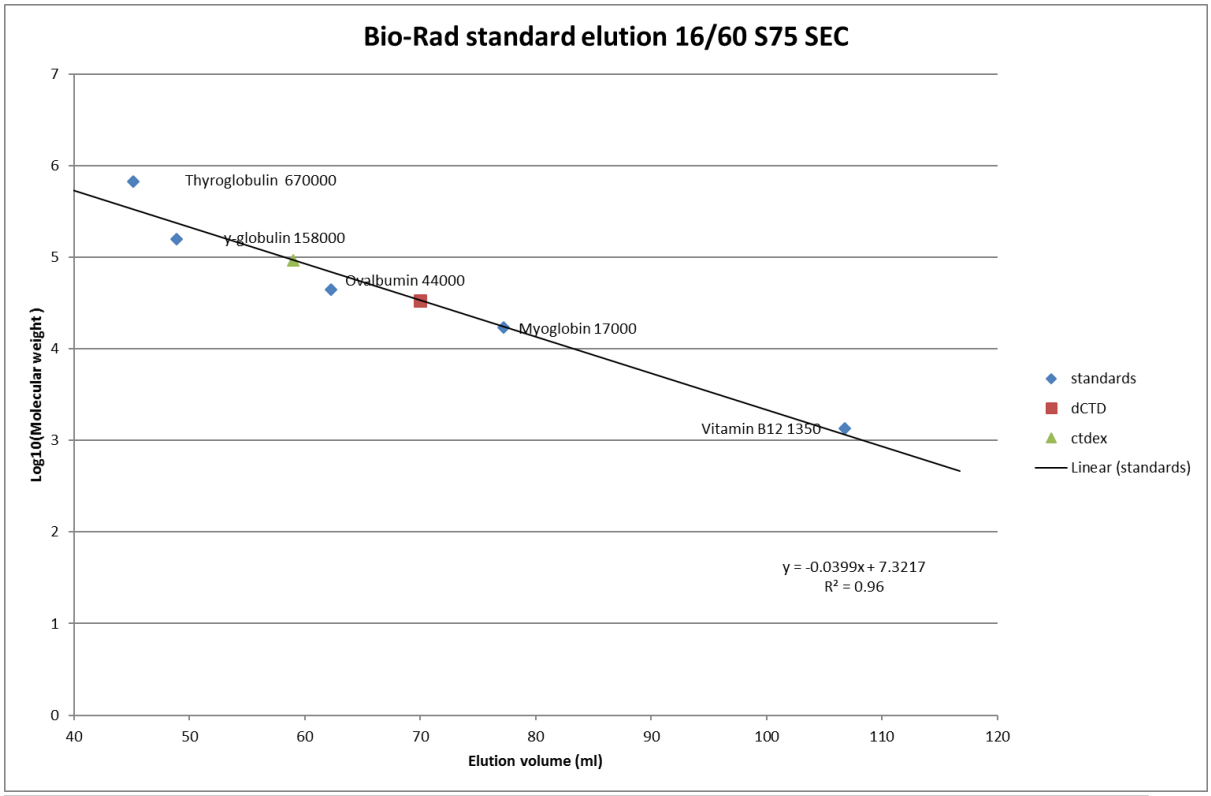
Supplementary Table 7.6: **list of antibodies.**

7.7 CtIP Twin Strep sequencing



Supplementary figure 7.7: **Sequence information for Twin-Strep tag insertion into CtIP$_{FL}$.** DNA sequence is shown with restriction sites displayed. Below is the CtIP$_{FL}$ open reading frame with protein translation and features shown. Underneath is the sequencing trace file confirming the sequence is correct. CtIP$_{FL}$ sequence was inserted into the plasmid by *Bam*HI restriction insertion followed by insertion of the Twin-Strep tag by *Sal*I restriction insertion. *Bst*BI site was included for identification of tag insertion.

7.8 dCTD and CTDex molecular weight standards



Supplementary figure 7.8: **Globular molecular weight standards on 16/60 Superdex 75 SEC column.** Protein samples dCTD, CTDex and bio-rad standards (Cat no. 1511901) were run sequentially on the same Superdex 75 16/60 SEC column with the same experimental setup. Elution Values for the standards were taken from the 280nm peaks using the Unicorn™ software and plotted in Excel verses their known $\log_{10}$ molecular weight. A straight line fit with equation shown was subsequently used to calculate the weight of dCTD and CTDex from their elution volumes to be 33.7 kDa and 92.8 kDa respectively.
The molecular weight of dCTD is higher than expected for a dimeric species, indicating either formation of a higher oligomeric state or a more elongated rod shape to the protein. CTDex also has a higher than expected weight for a dimeric species, further indicating part of the protein is elongated or formation of a higher oligomeric state.

7.9 Morpheus® crystallisation conditions

ID	Conc	Units	Salt	Conc	Units	Buffer	pH	Conc	Units	precipitant
A1	0.06	M	Divalents	0.1	M	Buffer 1	6.5	50	% v/v	Mix 1
A2	0.06	M	Divalents	0.1	M	Buffer 1	6.5	50	% v/v	Mix 2
A3	0.06	M	Divalents	0.1	M	Buffer 1	6.5	50	% v/v	Mix 3
A4	0.06	M	Divalents	0.1	M	Buffer 1	6.5	50	% v/v	Mix 4
A5	0.06	M	Divalents	0.1	M	Buffer 2	7.5	50	% v/v	Mix 1
A6	0.06	M	Divalents	0.1	M	Buffer 2	7.5	50	% v/v	Mix 2
A7	0.06	M	Divalents	0.1	M	Buffer 2	7.5	50	% v/v	Mix 3
A8	0.06	M	Divalents	0.1	M	Buffer 2	7.5	50	% v/v	Mix 4
A9	0.06	M	Divalents	0.1	M	Buffer 3	8.5	50	% v/v	Mix 1
A10	0.06	M	Divalents	0.1	M	Buffer 3	8.5	50	% v/v	Mix 2
A11	0.06	M	Divalents	0.1	M	Buffer 3	8.5	50	% v/v	Mix 3
A12	0.06	M	Divalents	0.1	M	Buffer 3	8.5	50	% v/v	Mix 4
B1	0.09	M	Halogens	0.1	M	Buffer 1	6.5	50	% v/v	Mix 1
B2	0.09	M	Halogens	0.1	M	Buffer 1	6.5	50	% v/v	Mix 2
B3	0.09	M	Halogens	0.1	M	Buffer 1	6.5	50	% v/v	Mix 3
B4	0.09	M	Halogens	0.1	M	Buffer 1	6.5	50	% v/v	Mix 4
B5	0.09	M	Halogens	0.1	M	Buffer 2	7.5	50	% v/v	Mix 1
B6	0.09	M	Halogens	0.1	M	Buffer 2	7.5	50	% v/v	Mix 2
B7	0.09	M	Halogens	0.1	M	Buffer 2	7.5	50	% v/v	Mix 3
B8	0.09	M	Halogens	0.1	M	Buffer 2	7.5	50	% v/v	Mix 4
B9	0.09	M	Halogens	0.1	M	Buffer 3	8.5	50	% v/v	Mix 1
B10	0.09	M	Halogens	0.1	M	Buffer 3	8.5	50	% v/v	Mix 2
B11	0.09	M	Halogens	0.1	M	Buffer 3	8.5	50	% v/v	Mix 3
B12	0.09	M	Halogens	0.1	M	Buffer 3	8.5	50	% v/v	Mix 4
C1	0.09	M	NPS	0.1	M	Buffer 1	6.5	50	% v/v	Mix 1
C2	0.09	M	NPS	0.1	M	Buffer 1	6.5	50	% v/v	Mix 2
C3	0.09	M	NPS	0.1	M	Buffer 1	6.5	50	% v/v	Mix 3
C4	0.09	M	NPS	0.1	M	Buffer 1	6.5	50	% v/v	Mix 4
C5	0.09	M	NPS	0.1	M	Buffer 2	7.5	50	% v/v	Mix 1
C6	0.09	M	NPS	0.1	M	Buffer 2	7.5	50	% v/v	Mix 2
C7	0.09	M	NPS	0.1	M	Buffer 2	7.5	50	% v/v	Mix 3
C8	0.09	M	NPS	0.1	M	Buffer 2	7.5	50	% v/v	Mix 4
C9	0.09	M	NPS	0.1	M	Buffer 3	8.5	50	% v/v	Mix 1
C10	0.09	M	NPS	0.1	M	Buffer 3	8.5	50	% v/v	Mix 2
C11	0.09	M	NPS	0.1	M	Buffer 3	8.5	50	% v/v	Mix 3
C12	0.09	M	NPS	0.1	M	Buffer 3	8.5	50	% v/v	Mix 4
D1	0.12	M	Alcohols	0.1	M	Buffer 1	6.5	50	% v/v	Mix 1
D2	0.12	M	Alcohols	0.1	M	Buffer 1	6.5	50	% v/v	Mix 2
D3	0.12	M	Alcohols	0.1	M	Buffer 1	6.5	50	% v/v	Mix 3
D4	0.12	M	Alcohols	0.1	M	Buffer 1	6.5	50	% v/v	Mix 4
D5	0.12	M	Alcohols	0.1	M	Buffer 2	7.5	50	% v/v	Mix 1
D6	0.12	M	Alcohols	0.1	M	Buffer 2	7.5	50	% v/v	Mix 2
D7	0.12	M	Alcohols	0.1	M	Buffer 2	7.5	50	% v/v	Mix 3
D8	0.12	M	Alcohols	0.1	M	Buffer 2	7.5	50	% v/v	Mix 4
D9	0.12	M	Alcohols	0.1	M	Buffer 3	8.5	50	% v/v	Mix 1
D10	0.12	M	Alcohols	0.1	M	Buffer 3	8.5	50	% v/v	Mix 2
D11	0.12	M	Alcohols	0.1	M	Buffer 3	8.5	50	% v/v	Mix 3
D12	0.12	M	Alcohols	0.1	M	Buffer 3	8.5	50	% v/v	Mix 4
E1	0.12	M	Ethylene glycols	0.1	M	Buffer 1	6.5	50	% v/v	Mix 1

ID	Conc	Units	Salt	Conc	Units	Buffer	pH	Conc	Units	precipitant
E2	0.12	M	Ethylene glycols	0.1	M	Buffer 1	6.5	50	% v/v	Mix 2
E3	0.12	M	Ethylene glycols	0.1	M	Buffer 1	6.5	50	% v/v	Mix 3
E4	0.12	M	Ethylene glycols	0.1	M	Buffer 1	6.5	50	% v/v	Mix 4
E5	0.12	M	Ethylene glycols	0.1	M	Buffer 2	7.5	50	% v/v	Mix 1
E6	0.12	M	Ethylene glycols	0.1	M	Buffer 2	7.5	50	% v/v	Mix 2
E7	0.12	M	Ethylene glycols	0.1	M	Buffer 2	7.5	50	% v/v	Mix 3
E8	0.12	M	Ethylene glycols	0.1	M	Buffer 2	7.5	50	% v/v	Mix 4
E9	0.12	M	Ethylene glycols	0.1	M	Buffer 3	8.5	50	% v/v	Mix 1
E10	0.12	M	Ethylene glycols	0.1	M	Buffer 3	8.5	50	% v/v	Mix 2
E11	0.12	M	Ethylene glycols	0.1	M	Buffer 3	8.5	50	% v/v	Mix 3
E12	0.12	M	Ethylene glycols	0.1	M	Buffer 3	8.5	50	% v/v	Mix 4
F1	0.12	M	Monosaccharides	0.1	M	Buffer 1	6.5	50	% v/v	Mix 1
F2	0.12	M	Monosaccharides	0.1	M	Buffer 1	6.5	50	% v/v	Mix 2
F3	0.12	M	Monosaccharides	0.1	M	Buffer 1	6.5	50	% v/v	Mix 3
F4	0.12	M	Monosaccharides	0.1	M	Buffer 1	6.5	50	% v/v	Mix 4
F5	0.12	M	Monosaccharides	0.1	M	Buffer 2	7.5	50	% v/v	Mix 1
F6	0.12	M	Monosaccharides	0.1	M	Buffer 2	7.5	50	% v/v	Mix 2
F7	0.12	M	Monosaccharides	0.1	M	Buffer 2	7.5	50	% v/v	Mix 3
F8	0.12	M	Monosaccharides	0.1	M	Buffer 2	7.5	50	% v/v	Mix 4
F9	0.12	M	Monosaccharides	0.1	M	Buffer 3	8.5	50	% v/v	Mix 1
F10	0.12	M	Monosaccharides	0.1	M	Buffer 3	8.5	50	% v/v	Mix 2
F11	0.12	M	Monosaccharides	0.1	M	Buffer 3	8.5	50	% v/v	Mix 3
F12	0.12	M	Monosaccharides	0.1	M	Buffer 3	8.5	50	% v/v	Mix 4
G1	0.1	M	Carboxylic acids	0.1	M	Buffer 1	6.5	50	% v/v	Mix 1
G2	0.1	M	Carboxylic acids	0.1	M	Buffer 1	6.5	50	% v/v	Mix 2
G3	0.1	M	Carboxylic acids	0.1	M	Buffer 1	6.5	50	% v/v	Mix 3
G4	0.1	M	Carboxylic acids	0.1	M	Buffer 1	6.5	50	% v/v	Mix 4
G5	0.1	M	Carboxylic acids	0.1	M	Buffer 2	7.5	50	% v/v	Mix 1
G6	0.1	M	Carboxylic acids	0.1	M	Buffer 2	7.5	50	% v/v	Mix 2
G7	0.1	M	Carboxylic acids	0.1	M	Buffer 2	7.5	50	% v/v	Mix 3
G8	0.1	M	Carboxylic acids	0.1	M	Buffer 2	7.5	50	% v/v	Mix 4
G9	0.1	M	Carboxylic acids	0.1	M	Buffer 3	8.5	50	% v/v	Mix 1
G10	0.1	M	Carboxylic acids	0.1	M	Buffer 3	8.5	50	% v/v	Mix 2
G11	0.1	M	Carboxylic acids	0.1	M	Buffer 3	8.5	50	% v/v	Mix 3
G12	0.1	M	Carboxylic acids	0.1	M	Buffer 3	8.5	50	% v/v	Mix 4
H1	0.1	M	Amino acids	0.1	M	Buffer 1	6.5	50	% v/v	Mix 1
H2	0.1	M	Amino acids	0.1	M	Buffer 1	6.5	50	% v/v	Mix 2
H3	0.1	M	Amino acids	0.1	M	Buffer 1	6.5	50	% v/v	Mix 3
H4	0.1	M	Amino acids	0.1	M	Buffer 1	6.5	50	% v/v	Mix 4
H5	0.1	M	Amino acids	0.1	M	Buffer 2	7.5	50	% v/v	Mix 1
H6	0.1	M	Amino acids	0.1	M	Buffer 2	7.5	50	% v/v	Mix 2
H7	0.1	M	Amino acids	0.1	M	Buffer 2	7.5	50	% v/v	Mix 3
H8	0.1	M	Amino acids	0.1	M	Buffer 2	7.5	50	% v/v	Mix 4
H9	0.1	M	Amino acids	0.1	M	Buffer 3	8.5	50	% v/v	Mix 1
H10	0.1	M	Amino acids	0.1	M	Buffer 3	8.5	50	% v/v	Mix 2
H11	0.1	M	Amino acids	0.1	M	Buffer 3	8.5	50	% v/v	Mix 3
H12	0.1	M	Amino acids	0.1	M	Buffer 3	8.5	50	% v/v	Mix 4

Supplementary figure 7.8-1: **condition mixes in Morpheus® commercial screen.**

'Salt'	Composition
Divalents	0.3M Magnesium chloride hexahydrate; 0.3M Calcium chloride dihydrate
Halogens	0.3M Sodium fluoride; 0.3M Sodium bromide; 0.3M Sodium iodide
NPS (Nitrate Phosphate Sulfate)	0.3M Sodium nitrate, 0.3 Sodium phosphate dibasic, 0.3M Ammonium sulfate
Alcohols	0.2M 1,6-Hexanediol; 0.2M 1-Butanol; 0.2M 1,2-Propanediol; 0.2M 2-Propanol; 0.2M 1,4-Butanediol; 0.2M 1,3-Propanediol
Ethylene glycols	0.3M Diethylene glycol; 0.3M Triethylene-glycol; 0.3M Tetraethylene glycol; 0.3M Pentaethylene glycol
Monosaccharides	0.2M D-Glucose; 0.2M D-Mannose; 0.2M D-Galactose; 0.2M L-Fucose; 0.2M D-Xylose; 0.2M N-Acetyl-D-Glucosamine
Carboxylic acids	0.2M Sodium formate; 0.2M Ammonium acetate; 0.2M Sodium citrate tribasic dihydrate; 0.2M Sodium potassium tartrate tetrahydrate; 0.2M Sodium oxamate
Amino acids	0.2M L-Na-Glutamate; 0.2M Alanine (racemic); 0.2M Glycine; 0.2M Lysine HCl (racemic); 0.2M Serine (racemic)

Supplementary figure 7.8-2: **Mixes of additives in Morpheus® commercial screen.**

Mix name	Conc.	pH @ 20°C	Composition
Buffer System 1	1.0 M	6.5	Imidazole; MES monohydrate (acid)
Buffer System 2	1.0 M	7.5	Sodium HEPES; MOPS (acid)
Buffer System 3	1.0 M	8.5	Tris (base); BICINE

Supplementary figure 7.8-3: **Buffer systems in Morpheus® commercial screen**

Mix name	Composition
Precipitant Mix 1	40% v/v PEG 500* MME; 20 % w/v PEG 20000
Precipitant Mix 2	40% v/v Ethylene glycol; 20 % w/v PEG 8000
Precipitant Mix 3	40% v/v Glycerol; 20% w/v PEG 4000
Precipitant Mix 4	25% v/v MPD; 25% PEG 1000; 25% w/v PEG 3350

Supplementary figure 7.8-4**: Mixes of Precipitants in Morpheus® commercial screen.**

Appendix 2, Primers

N-terminal domain

Residue	Primer	Enzyme
Fwd CtIP 1	CATGCCATGGC AACATCTCGGGAAGCAGCTG	*NcoI*
Fwd CtIP 28	CATGCCATGGGCAAAGAATGTCATGATAGAGAAGTACAA	*NcoI*
Fwd CtIP 29	CATGCCATGGGCGAATGTCATGATAGAGAAGTACAAGG	*NcoI*
Fwd CtIP 30	CATGCCATGGGCTGTCATGATAGAGAAGTACAAGGTTTAC	*NcoI*
Fwd CtIP 31	CATGCCATGGGCCATGATAGAGAAGTACAAGGTTTACAAG	*NcoI*
Fwd CtIP 32	CATGCCATGGGCGATAGAGAAGTACAAGGTTTACAAGTAAAA G	*NcoI*
Fwd CtIP 33	CATGCCATGGGCAGAGAAGTACAAGGTTTACAAGTAAAAGT	*NcoI*
Fwd CtIP 34	CATGCCATGGGCGAAGTACAAGGTTTACAAGTAAAAGTAACC	*NcoI*
Fwd CtIP 35	CATGCCATGGGCGTACAAGGTTTACAAGTAAAAGTAACCAAGC	*NcoI*
Fwd CtIP 36	CATGCCATGGGCCAAGGTTTACAAGTAAAAGTAACCAAGCTAA AAC	*NcoI*
Fwd CtIP 37	CATGCCATGGGCGGTTTACAAGTAAAAGTAACCAAGCTAAAAC AG	*NcoI*
Fwd CtIP 38	CATGCCATGGGCTTACAAGTAAAAGTAACCAAGCTAAAACAGG A	*NcoI*
Rev CtIP 88 Strep	CGCTCGAGCTACTTCTCAAACTGAGGGTGGCTCCATAAGCCTG C TCTTAACCG	*XhoI*
Fwd CtIP 93	CATGCCATGGGCGCAGTAACTGAAGAACATATGC	*NcoI*
Rev CtIP 145	CGCTCGAGCTAATCATTCTCAATTTTCTGCTG	*XhoI*
Rev CtIP 145 Strep	CGCTCGAGCTACTTCTCAAACTGAGGGTGGCTCCAATCATTCT CA ATTTTCTGCTG	*XhoI*

N-terminal domain mutation primers

Mutation	Primer
Cluster 1 KKQ Fwd	GAACATATGCGGGCCGCAGCGCAAGAGTTTGA
Cluster 1 KKQ Rev	CTGAAGAACATATGCGGGCCGCAGCGCAAGAGTTTGAAAATATCC
Cluster 2 KQE Fwd	TTACAAGTAAAAGTAACCAAGCTAGCAGCGGCCCGAATCTTAGAT GCACAAA
Cluster 2 KQE Rev	GTAACCAAGCTAGCAGCGGCCCGAATCTTAGATG
Cluster 3 EQK Fwd	GAATACTCTACAGGCCGCGAATGCAAAGCTTTCTG
Cluster 3 EQK Rev	CAGAAAGCTTTGCATTCGCGGCCTGTAGAGTATTC

C-terminal domain

Primer	Sequence	Enzyme
Fwd CtIP 694	CTAGTACCATGGGCTTGGTTAGTGAAACCGTTC	*Nco*I
Rev CtIP 897 strep	CGCTCGAGCTACTTCTCAAACTGAGGGTGGCTCCACGCTTTC TGCTTGACTTT	*Xho*I
Rev CtIP 897	CATGAAGCGGCCGCCTCGAGTTACTATGTCTTCTGCTCCTTG CC	*Xho*I

C-terminal domain mutation primers

Primer	Sequence
769_FWD	CTAGTACCATGGACAAAGTCAAGCAGAAAGC
897_REV	CATGAAGCGGCCGCCTCGAGTTACTATGTCTTCTGCTCCTTGCC
C795A_FWD	GCTTGCAAAATTTTCCTGCGATTGAGGTGGTTCGG
C795A_REV	CCGAACCACCTCAATCGCAGGAAAATTTTGCAAGC
H811A_FWD	GAGAAGAAAACTGCTTGGGGCCACGTGTAAGGAATGTGAAA
H811A_REV	TTTCACATTCCTTACACGTGGCCCCAAGCAGTTTTCTTCTC
C813A_FWD	CTGCTTGGGCACACGGCGAAGGAATGTGAAATTTATTATGC
C813A_REV	GCATAATAAATTTCACATTCCTTCGCCGTGTGCCCAAGCAG
C816A_FWD	GCACACGTGTAAGGAAGCGGAAATTTATTATGCAG
C816A_REV	CTGCATAATAAATTTCCGCTTCCTTACACGTGTGC
C835A_FWD	GAGAAAAGAAATTGGCTTCCGCGTCAAGACACCGATTCCG
C835A_REV	CGGAATCGGTGTCTTGACGCGGAAGCCAATTTCTTTTCTC
H838A_FWD	GGCTTCCTGCTCAAGAGCCCGATTCCGCTACATTC
H838A_REV	GAATGTAGCGGAATCGGGCTCTTGAGCAGGAAGCC
C862A_FWD	CCACTCAGACTGCGATGGAAAGAGG
C862A_REV	CCTCTTTCCATCGCAGTCTGAGTGG
C875A_FWD	TATTAAGGAAGATCTTGATCCTGCGCCTCGTCCAAAAAGACG
C875A_REV	CGTCTTTTTGGACGAGGCGCAGGATCAAGATCTTCCTTAATA
T847E_FWD	GCTACATTCCACCCAACGAACCAGAGAATTTTTGGGA
T847E_REV	TCCCAAAAATTCTCTGGTTCGTTGGGTGGAATGTAGC

H811A&C813A_FWD	GAGAAGAAAACTGCTTGGGGCCACGGCGAAGGAATGTGAAATTTATTATGC
H811A&C813A_REV	GCATAATAAATTTCACATTCCTTCGCCGTGGCCCCAAGCAGTTTTCTTCTC
H811A&C813A&C816A_FWD	GAGAAGAAAACTGCTTGGGGCCACGGCGAAGGAAGCGGAAATTTATTATGCAG
H811A&C813A&C816A_REV	CTGCATAATAAATTTCCGCTTCCTTCGCCGTGGCCCCAAGCAGTTTTCTTCTC
H811A&C816A_FWD	GAGAAGAAAACTGCTTGGGGCCACGTGTAAGGAAGCGGAAATTTATTATGCAG
H811A&C816A_REV	CTGCATAATAAATTTCCGCTTCCTTACACGTGGCCCCAAGCAGTTTTCTTCTC
C813A&C816A_FWD	GAAAACTGCTTGGGCACACGGCGAAGGAAGCGGAAATTTATTATGCAG
C813A&C816A_REV	CTGCATAATAAATTTCCGCTTCCTTCGCCGTGTGCCCAAGCAGTTTT
C835A&H838A_FWD	GAAAAGAAATTGGCTTCCGCGTCAAGAGCCCGATTCCGC
C835A&H838A_REV	GCGGAATCGGGCTCTTGACGCGGAAGCCAATTTCTTTTC

CtIP Full-length

Primer	Sequence	Enzyme
CtIP *E. coli* Fwd	CATGACATGTCTAACATCTCGGGAAGCAGCTG	*Sal*I-HF
CtIP *E. coli* Rev	CGGTCGACCTATGTCTTCTGCTCCTTGCC	*Nco*I-HF
CtIP Fwd	CATATTGGATCCGGGTCGACCGCCACCATGAACATCTCGGGAAGC	*Bam*HI*Sal*I
CtIP Rev	ACATTTGCGGCCGCTTATGTCTTCTGCTCCTTGC	*Not*I
Twin-Strep Fwd	CATGATGTCGACCGCCACCATGAGCG	*Sal*I
Twin-Strep Rev	TTCATAGTCGACCCTTGAAAATACAGGTTCTCCTT	*Sal*I
L27E-Fwd	GAATGTCATGATAGAGAAGTACAAGGTGCAGCACAAGTAAAAGTAACCAACGTAAAAC	*N/A*
L27E-Rev	GTTTTACGTTGGTTACTTTTACTTGTGCTGCACCTTGTACTTCTCTATCATGACATTC	*N/A*

Sequencing Primers

Primer	Sequence
CtIP internal 1	TTCTAGTAGCTGACACTTATGACCAAAG
CtIP internal 2	CTCCCTTCTGCTCTTGTTTC
MalE	GGTCGTCAGACTGTCGATGAAGCC
T7 terminator	GCTAGTTATTGCTCAGCGG
SV40 Rev	CCTCTACAAATGTGGTATGG
PolH Fwd	AAATGATAACCATCTCGC
CMV fwd	CGCAAATGGGCGGTAGGCGTG

Appendix 3, Construct abbreviations

CtIP N-terminal domain

CtIP residues 1-897	$CtIP_{FL}$
CtIP residues 1-897 L27E	$CtIP_{L27E}$
CtIP residues 1-145	NTD_{1-145}
CtIP residues 18-145	NTD_{18-145}
CtIP residues 18-145 N-terminal MBP fusion	MBP-NTD_{18-145}
CtIP residues 28-145	NTD_{28-145}
CtIP residues 29-145	NTD_{29-145}
CtIP residues 30-145	NTD_{30-145}
CtIP residues 31-145	NTD_{31-145}
CtIP residues 32-145	NTD_{32-145}
CtIP residues 33-145	NTD_{33-145}
CtIP residues 34-145	NTD_{34-145}
CtIP residues 35-145	NTD_{35-145}
CtIP residues 36-145	NTD_{36-145}
CtIP residues 37-145	NTD_{37-145}
CtIP residues 38-145	NTD_{38-145}
CtIP residues 28-145 C-terminal Strep tag	NTD_{28-145}-Strep
CtIP residues 29-145 C-terminal Strep tag	NTD_{29-145}-Strep
CtIP residues 30-145 C-terminal Strep tag	NTD_{30-145}-Strep
CtIP residues 31-145 C-terminal Strep tag	NTD_{31-145}-Strep
CtIP residues 32-145 C-terminal Strep tag	NTD_{32-145}-Strep
CtIP residues 33-145 C-terminal Strep tag	NTD_{33-145}-Strep
CtIP residues 34-145 C-terminal Strep tag	NTD_{34-145}-Strep
CtIP residues 35-145 C-terminal Strep tag	NTD_{35-145}-Strep
CtIP residues 36-145 C-terminal Strep tag	NTD_{36-145}-Strep
CtIP residues 37-145 C-terminal Strep tag	NTD_{37-145}-Strep
CtIP residues 38-145 C-terminal Strep tag	NTD_{38-145}-Strep
CtIP residues 30-145 N-terminal MBP fusion C-terminal Strep tag	MBP-NTD_{30-145}-Strep
CtIP residues 31-145 C89A C92A C-terminal strep tag	NTD_{31-145}DM-Strep
CtIP residues 31-145 K107A K108A Q109A - Cluster 1 - C-terminal strep tag	NTD_{31-145}C1-Strep
CtIP residues 31-145 K52A Q53A E54A - Cluster 2 - C-terminal strep tag	NTD_{31-145}C2-Strep
CtIP residues 31-145 E135A Q136A K138A - Cluster 3 - C-terminal strep tag	NTD_{31-145}C3-Strep
CtIP residues 31-88 C-terminal strep tag	NTD_{31-88}-Strep
CtIP residues 93-145 C-terminal strep tag	NTD_{93-145}-Strep

CtIP C-terminal domain

CtIP residues 769-897	CtIP-CTD
MBP fused CtIP residues 769-897	MBP-CTD
CtIP residues 694-775	dCTD
CtIP residues 694-897	CTDex
CtIP residues 769-897 C813A C816A	CTD_{DM}
CtIP residues 769-897 H795A	CTD_{H795A}
CtIP residues 769-897 H811A	CTD_{H811A}
CtIP residues 769-897 C813A	CTD_{C813A}
CtIP residues 769-897 C816A	CTD_{C816A}
CtIP residues 769-897 C835A	CTD_{C835A}
CtIP residues 769-897 H383A	CTD_{H383A}
CtIP residues 769-897 C862A	CTD_{C862A}
CtIP residues 769-897 C875A	CTD_{C875A}
CtIP residues 769-897 H795A, H811A, C835A, H838A, C862A, C875A	CTD_{ABC}
CtIP residues 769-897 T847E	CTD_{T847E}

And1

And1 residues 1-1129	$And1_{FL}$
And1 residues 826-1129	$And1_{826-1129}$
And1 residues 316-1129	$And1_{316-1129}$
And1 residues 1-316	$And1_{1-316}$
And1 residues 1-826	$And1_{1-826}$
And1 residues 316-826	$And1_{316-826}$

Bibliography

Abbott, P.J., Saffhill, R.O.Y., 1979. DNA synthesis with methylated Poly dC-dG templates. Evidence for a competitive nature to miscoding by O6-methylguanine. Elsevier/north-holl. Biomed. 562, 51–61.

Abe, T., Kawasumi, R., Giannattasio, M., Dusi, S., Yoshimoto, Y., Miyata, K., Umemura, K., Hirota, K., Branzei, D., 2018. AND-1 fork protection function prevents fork resection and is essential for proliferation. Nat. Commun. 9. doi:10.1038/s41467-018-05586-7

Adamo, A., Collis, S.J., Adelman, C. a., Silva, N., Horejsi, Z., Ward, J.D., Martinez-Perez, E., Boulton, S.J., La Volpe, A., 2010. Preventing Nonhomologous End Joining Suppresses DNA Repair Defects of Fanconi Anemia. Mol. Cell 39, 25–35. doi:10.1016/j.molcel.2010.06.026

Adams, P.D., Afonine, P. V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., McCoy, A.J., Moriarty, N.W., Oeffner, R., Read, R.J., Richardson, D.C., Richardson, J.S., Terwilliger, T.C., Zwart, P.H., 2010. PHENIX: A comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. Sect. D Biol. Crystallogr. 66, 213–221. doi:10.1107/S0907444909052925

Adebanke F. Fagbemi., Orelli, B., Schärer, O.D., 2011. Regulation of endonuclease activity in human nucleotide excision repair. DNA repair 10, 722–729. doi:10.1021/nl061786n.Core-Shell

Aguilera, A., Gómez-González, B., 2008. Genome instability: A mechanistic view of its causes and consequences. Nat. Rev. Genet. 9, 204–217. doi:10.1038/nrg2268

Ahnesorg, P., Smith, P., Jackson, S.P., 2006. XLF interacts with the XRCC4-DNA Ligase IV complex to promote DNA nonhomologous end-joining. Cell 124, 301–313. doi:10.1016/j.cell.2005.12.031

Ahrland, S., Herman, R.G., 1975. Spectrophotometric Determination of Manganese(II) and Zinc(II) with 4-(2-Pyridylazo)Resorcinol (PAR). Anal. Chem. 47, 2422–2426. doi:10.1021/ac60364a028

Alexander, J.L., Orr-weaver, T.L., 2016. Replication fork instability and the consequences of fork collisions from rereplication 2241–2252. doi:10.1101/gad.288142.116.

Althubiti, M., Lezina, L., Carrera, S., Jukes-Jones, R., Giblett, S.M., Antonov, a., Barlev, N., Saldanha, G.S., Pritchard, C. a., Cain, K., MacIp, S., 2014. Characterization of novel markers of senescence and their prognostic potential in cancer. Cell Death Dis. 5, e1528-10. doi:10.1038/cddis.2014.489

Anand, R., Ranjha, L., Cannavo, E., Cejka, P., 2016. Phosphorylated CtIP Functions as a Co-factor of the MRE11-RAD50-NBS1 Endonuclease in DNA End Resection. Mol. Cell 64, 940–950. doi:10.1016/j.molcel.2016.10.017

Andres, S.N., Appel, C.D., Westmoreland, J.W., Williams, J.S., Nguyen, Y., Robertson, P.D., Resnick, M. a, Williams, R.S., 2015. Tetrameric Ctp1 coordinates DNA binding and DNA bridging in DNA double-strand-break repair. Nat. Struct. Mol. Biol. 2. doi:10.1038/nsmb.2945

Ashcroft, M., Kubbutat, M.H.G., Vousden, K.H., 1999. Regulation of p53 function and stability by phosphorylation. Mol. Cell. Biol. 19, 1751–1758. doi:10.1128/MCB.19.3.1751

Audebert, M., Salles, B., Weinfeld, M., Calsou, P., 2006. Involvement of polynucleotide kinase in a poly(ADP-ribose) polymerase-1-dependent DNA double-strand breaks rejoining pathway. J. Mol. Biol. 356, 257–265. doi:10.1016/j.jmb.2005.11.028

Audeh, M.W., Carmichael, J., Penson, R.T., Friedlander, M., Powell, B., Bell-McGuinn, K.M., Scott, C., Weitzel, J.N., Oaknin, A., Loman, N., Lu, K., Schmutzler, R.K., Matulonis, U., Wickens, M., Tutt, A., 2010. Oral poly(ADP-ribose) polymerase inhibitor olaparib in patients with BRCA1 or BRCA2 mutations and recurrent ovarian cancer: a proof-of-concept trial. Lancet 376, 245–251. doi:10.1016/S0140-6736(10)60893-8

Baber-Furnari, B. a, Rhind, N., Boddy, M.N., Shanahan, P., Lopez-Girona, a, Russell, P., 2000. Regulation of mitotic inhibitor Mik1 helps to enforce the DNA damage checkpoint. Mol. Biol. Cell 11, 1–11. doi:10.1091/MBC.11.1.1

Backliwal, G., Hildinger, M., Kuettel, I., Delegrange, F., Hacker, D.L., Wurm, F.M., 2008. Valproic acid:

A viable alternative to sodium butyrate for enhancing protein expression in mammalian cell cultures. Biotechnol. Bioeng. 101, 182–189. doi:10.1002/bit.21882
Bakkenist, C.J., Kastan, M.B., 2003. DNA damage activates ATM through intermolecular autophosphorylation and dimer dissociation. Nature. doi:10.1038/nature01368
Banin, S., Moyal, L., Shieh, S.-Y., Taya, Y., Anderson, C.W., Chessa, L., Smorodinsky, N.I., Prives, C., Reiss, Y., Shiloh, Y., Ziv, Y., 1998. Enhanced Phosphorylation of p53 by ATM in Response to DNA Damage. Science (80-.). 281, 1674 LP-1677.
Barnum, K.J., O'Connell, M.J., 2014. Cell Cycle Regulation by Checkpoints. Methods Mol. Biol. 1170, 29–40. doi:10.1007/978-1-4939-0888-2_2
Bartek, J., Lukas, C., Lukas, J., 2004. Checking on DNA damage in S phase. Nat. Rev. Mol. Cell Biol. 5, 792–804. doi:10.1038/nrm1493
Barton, O., Naumann, S.C., Diemer-Biehs, R., Künzel, J., Steinlage, M., Conrad, S., Makharashvili, N., Wang, J., Feng, L., Lopez, B.S., Paull, T.T., Chen, J., Jeggo, P. a, Löbrich, M., 2014. Polo-like kinase 3 regulates CtIP during DNA double-strand break repair in G1. J. Cell Biol. 206, 877–94. doi:10.1083/jcb.201401146
Batchelor, E., Loewer, A., Lahav, G., 2009. The ups and downs of p53: understanding protein dynamics in single cells. Nat. Rev. Cancer 9, 371.
Baudat, F., De Massy, B., 2007. Regulating double-stranded DNA break repair towards crossover or non-crossover during mammalian meiosis. Chromosom. Res. 15, 565–577. doi:10.1007/s10577-007-1140-3
Bębenek, A., Ziuzia-Graczyk, I., 2018. Fidelity of DNA replication—a matter of proofreading. Curr. Genet. 64, 985–996. doi:10.1007/s00294-018-0820-1
Beck, H., Nähse, V., Larsen, M.S.Y., Groth, P., Clancy, T., Lees, M., Jørgensen, M., Helleday, T., Syljuåsen, R.G., Sørensen, C.S., 2010. Regulators of cyclin-dependent kinases are crucial for maintaining genome integrity in S phase. J. Cell Biol. 188, 629–638. doi:10.1083/jcb.200905059
Bennardo, N., Cheng, A., Huang, N., Stark, J.M., 2008. Alternative-NHEJ is a mechanistically distinct pathway of mammalian chromosome break repair. PLoS Genet. 4, e1000110. doi:10.1371/journal.pgen.1000110
Bertoli, C., Herlihy, A.E., Pennycook, B.R., Kriston-Vizi, J., De Bruin, R. a. M., 2016. Sustained E2F-Dependent Transcription Is a Key Mechanism to Prevent Replication-Stress-Induced DNA Damage. Cell Rep. 15, 1412–1422. doi:10.1016/j.celrep.2016.04.036
Besson, A., Dowdy, S.F., Roberts, J.M., 2008. CDK Inhibitors: Cell Cycle Regulators and Beyond. Dev. Cell 14, 159–169. doi:10.1016/j.devcel.2008.01.013
Blackford, A.N., Jackson, S.P., 2017. ATM, ATR, and DNA-PK: The Trinity at the Heart of the DNA Damage Response. Mol. Cell 66, 801–817. doi:10.1016/j.molcel.2017.05.015
Blackford, A.N., Nieminuszczy, J., Schwab, R. a., Galanty, Y., Jackson, S.P., Niedzwiedz, W., 2015. TopBP1 Interacts with BLM to Maintain Genome Stability but Is Dispensable for Preventing BLM Degradation. Mol. Cell 57, 1133–1141. doi:10.1016/j.molcel.2015.02.012
Bøe, C. a., Håland, T.W., Boye, E., Syljuåsen, R.G., Grallert, B., 2018. A novel role for ATR/Rad3 in G1 phase. Sci. Rep. 8, 1–12. doi:10.1038/s41598-018-25238-6
Borja-Cacho, D., Matthews, J., 2008. Direct reversal of DNA Alkylation damage. NIH Public Access 6, 2166–2171. doi:10.1021/nl061786n.Core-Shell
Boswell-Casteel, R.C., Johnson, J.M., Duggan, K.D., Tsutsui, Y., Hays, F.A., 2015. Overproduction and biophysical characterization of human HSP70 proteins. Protein Expr. Purif. 106, 57–65. doi:10.1016/j.pep.2014.09.013
Bouwman, P., Jonkers, J., 2012. The effects of deregulated DNA damage signalling on cancer chemotherapy response and resistance. Nat. Rev. Cancer 12, 587–598. doi:10.1038/nrc3342
Brown, E.., Baltimore, D., 2000. ATR disruption leads to chromosomal fragmentation and early embryonic lethality. 397–402. doi:10.1080/00222938409487407
Brown, E.J., Baltimore, D., 2003. Essential and dispensable roles of ATR in cell cycle arrest and genome maintenance. Genes Dev. 17, 615–628. doi:10.1101/gad.1067403

Brown, P.H., Balbo, A., Schuck, P., 2008. Characterizing Protein-Protein Interactions by Sedimentation Velocity Analytical Ultracentrifugation. Curr. Protoc. Immunol. 81, 18.15.1-18.15.39. doi:10.1002/0471142735.im1815s81

Bunting, S.F., Callén, E., Wong, N., Chen, H.T., Polato, F., Gunn, A., Bothmer, A., Feldhahn, N., Fernandez-Capetillo, O., Cao, L., Xu, X., Deng, C.X., Finkel, T., Nussenzweig, M., Stark, J.M., Nussenzweig, A., 2010. 53BP1 inhibits homologous recombination in brca1-deficient cells by blocking resection of DNA breaks. Cell 141, 243–254. doi:10.1016/j.cell.2010.03.012

Bunz, F., Dutriaux, A., Lengauer, C., Waldman, T., Zhou, S., Brown, J.P., Sedivy, J.M., Kinzler, K.W., Vogelstein, B., 1998. Requirement for p53 and p21 to Sustain G2 Arrest After DNA Damage. Science (80-.). 282, 1497 LP-1501.

Byun, T.S., Pacek, M., Yee, M.C., Walter, J.C., Cimprich, K. a., 2005. Functional uncoupling of MCM helicase and DNA polymerase activities activates the ATR-dependent checkpoint. Genes Dev. 19, 1040–1052. doi:10.1101/gad.1301205

Canman, C.E., Lim, D.-S., Cimprich, K.A., Taya, Y., Tamai, K., Sakaguchi, K., Appella, E., Kastan, M.B., Siliciano, J.D., 1998. Activation of the ATM Kinase by Ionizing Radiation and Phosphorylation of p53. Science (80-.). 281, 1677 LP-1679.

Cannavo, E., Cejka, P., 2014. Sae2 promotes dsDNA endonuclease activity within Mre11–Rad50–Xrs2 to resect DNA breaks. Nature. doi:10.1038/nature13771

Cannavo, E., Johnson, D., Andres, S.N., Kissling, V.M., Reinert, J.K., Garcia, V., Erie, D.A., Hess, D., Thomä, N.H., Enchev, R.I., Peter, M., Williams, R.S., Neale, M.J., Cejka, P., 2018. Regulatory control of DNA end resection by Sae2 phosphorylation. Nat. Commun. 1–14. doi:10.1038/s41467-018-06417-5

Cao, L., Xu, X., Bunting, S.F., Liu, J., Wang, R.H., Cao, L.L., Wu, J.J., Peng, T.N., Chen, J., Nussenzweig, A., Deng, C.X., Finkel, T., 2009. A Selective Requirement for 53BP1 in the Biological Response to Genomic Instability Induced by Brca1 Deficiency. Mol. Cell 35, 534–541. doi:10.1016/j.molcel.2009.06.037

Ceccaldi, R., Liu, J.C., Amunugama, R., Hajdu, I., Primack, B., Petalcorin, M.I.R., O'Connor, K.W., Konstantinopoulos, P. a., Elledge, S.J., Boulton, S.J., Yusufzai, T., D'Andrea, A.D., 2015. Homologous-recombination-deficient tumours are dependent on Polθ -mediated repair. Nature 518, 258–262. doi:10.1038/nature14184

Chang, H.H.Y., Pannunzio, N.R., Adachi, N., Lieber, M.R., 2017. Non-homologous DNA end joining and alternative pathways to double-strand break repair. Nat. Rev. Mol. Cell Biol. 18, 495–506. doi:10.1038/nrm.2017.48

Chanut, P., Britton, S., Coates, J., Jackson, S.P., Calsou, P., 2016. Coordinated nuclease activities counteract Ku at single-ended DNA double-strand breaks. Nat. Commun. 7, 12889. doi:10.1038/ncomms12889

Chapman, J.R., Taylor, M.R.G., Boulton, S.J., 2012. Playing the End Game: DNA Double-Strand Break Repair Pathway Choice. Mol. Cell 47, 497–510. doi:10.1016/j.molcel.2012.07.029

Chappell, C., Hanakahi, L. a., Karimi-Busheri, F., Weinfeld, M., West, S.C., 2002. Involvement of human polynucleotide kinase in double-strand break repair by non-homologous end joining. EMBO J. 21, 2827–2832. doi:10.1093/emboj/21.11.2827

Charpentier, M., Khedher, a. H.Y., Menoret, S., Brion, a., Lamribet, K., Dardillac, E., Boix, C., Perrouault, L., Tesson, L., Geny, S., De Cian, a., Itier, J.M., Anegon, I., Lopez, B., Giovannangeli, C., Concordet, J.P., 2018. CtIP fusion to Cas9 enhances transgene integration by homology-dependent repair. Nat. Commun. 9, 1–11. doi:10.1038/s41467-018-03475-7

Chen, L., Nievera, C.J., Lee, A.Y.-L., Wu, X., 2008. Cell Cycle-dependent Complex Formation of BRCA1·CtIP·MRN Is Important for DNA Double-strand Break Repair. J. Biol. Chem. 283, 7713–7720. doi:10.1074/jbc.M710245200

Chen, M., Ryan, C.E., Piwnica-worms, H., 2003. Chk1 Kinase Negatively Regulates Mitotic Function.pdf 23, 7488–7497. doi:10.1128/MCB.23.21.7488

Chen, P.-L., Liu, F., Cai, S., Lin, X., Li, A., Chen, Y., Gu, B., Lee, E.Y.-H.P., Lee, W.-H., 2005. Inactivation

of CtIP leads to early embryonic lethality mediated by G1 restraint and to tumorigenesis by haploid insufficiency. Mol. Cell. Biol. 25, 3535–42. doi:10.1128/MCB.25.9.3535-3542.2005
Chen, Y., Liu, H., Zhang, H., Sun, C., Hu, Z., Tian, Q., Peng, C., Jiang, P., Hua, H., Li, X., Pei, H., 2017. And-1 coordinates with CtIP for efficient homologous recombination and DNA damage checkpoint maintenance. Nucleic Acids Res. 45, 2516–2530. doi:10.1093/nar/gkw1212
Cheung-Ong, K., Giaever, G., Nislow, C., 2013. DNA-damaging agents in cancer chemotherapy: Serendipity and chemical biology. Chem. Biol. 20, 648–659. doi:10.1016/j.chembiol.2013.04.007
Chinnadurai, G., 2006. CtIP, a candidate tumor susceptibility gene is a team player with luminaries. Biochim. Biophys. Acta - Rev. Cancer 1765, 67–73. doi:10.1016/j.bbcan.2005.09.002
Chrzanowska, K.H., Gregorek, H., Dembowska-Bagińska, B., Kalina, M. a., Digweed, M., 2012. Nijmegen breakage syndrome (NBS). Orphanet J. Rare Dis. 7, 1–19. doi:10.1186/1750-1172-7-1322373003
Ciccia, A., Elledge, S.J., 2011. The DNA Damage Response : Making it safe to play with knives. mol cell 40, 179–204. doi:10.1016/j.molcel.2010.09.019.The
Cimprich, K.A., Shin, T.B., Keith, C.T., Schreiber, S.L., 1996. cDNA cloning and gene mapping of a candidate human cell cycle checkpoint protein. Proc. Natl. Acad. Sci. U. S. A. 93, 2850–2855.
Clamp, M., Cuff, J., Searle, S.M., Barton, G.J., 2004. The Jalview Java alignment editor. Bioinformatics 20, 426–427. doi:10.1093/bioinformatics/btg430
Clancy, S., 2008. DNA damage & repair: Mechanisms for maintaining DNA integrity, Nature Education.
Clerici, M., Mantiero, D., Lucchini, G., Longhese, M.P., 2005. The Saccharomyces cerevisiae Sae2 protein promotes resection and bridging of double strand break ends. J. Biol. Chem. 280, 38631–38638. doi:10.1074/jbc.M508339200
Cole, J.L., Lary, J.W., P. Moody, T., Laue, T.M.B.T.-M. in C.B., 2008. Analytical Ultracentrifugation: Sedimentation Velocity and Sedimentation Equilibrium, in: Biophysical Tools for Biologists, Volume One: In Vitro Techniques. Academic Press, pp. 143–179. doi:https://doi.org/10.1016/S0091-679X(07)84006-4
Coleman, T.R., Dunphy, W.G., 1994. cdc2 regulatory factors. Curr. Opin. Cell Biol. 6, 877–882. doi:10.1016/0955-0674(94)90060-4
Collaborative Computational Project, 1994. The CCP4 suite: programs for protein crystallography. Acta Crystallogr. Sect. D D50, 760–763.
Cooper, D.R., Boczek, T., Grelewska, K., Pinkowska, M., Sikorska, M., Zawadzki, M., Derewenda, Z., 2007. Protein crystallization by surface entropy reduction: Optimization of the SER strategy. Acta Crystallogr. Sect. D Biol. Crystallogr. 63, 636–645. doi:10.1107/S0907444907010931
Cooper, G.M., 2000. A Molecular Approach. 2nd edition. Sinauer Associates, Sunderland (MA).
Cortez, D., Guntuku, S., Qin, J., 2001. ATR and ATRIP : Partners in Checkpoint Signaling 294, 1713–1717. doi:10.1126/science.1065521
Cortez, D., Wang, Y., Qin, J., Elledge, S.J., 1999. Requirement of ATM-Dependent Phosphorylation of Brca1 in the DNA Damage Response to Double-Strand Breaks. Science (80-.). 286, 1162 LP-1166.
Crick, F.H.C., 1974. The double helix, a personal view. Nature 248, 766–769.
Crown, K.N., Savytskyy, O.P., Malik, S.-B., Logsdon, J., Williams, R.S., Tainer, J.A., Zolan, M.E., 2013. A Mutation in the FHA Domain of Coprinus cinereus Nbs1 Leads to Spo11-Independent Meiotic Recombination and Chromosome Segregation. G3 Genes|Genomes|Genetics 3, 1927 LP-1943.
Cruz-García, A., López-Saavedra, A., Huertas, P., 2014. BRCA1 Accelerates CtIP-Mediated DNA-End Resection. Cell Rep. 451–459. doi:10.1016/j.celrep.2014.08.076
Culurgioni, S., Muñoz, I.G., Palacios, A., Redondo, P., Blanco, F.J., Montoya, G., 2010. Crystallization and preliminary X-ray diffraction analysis of the dimerization domain of the tumour suppressor ING4. Acta Crystallogr. Sect. F 66, 567–570. doi:10.1107/S1744309110010080

Curtin, N.J., 2012. DNA repair dysregulation from cancer driver to therapeutic target. Nat. Rev. Cancer 12, 801.
D'Amours, D., Jackson, S.P., 2002. The Mre11 complex: at the crossroads of dna repair and checkpoint signalling. Nat. Rev. Mol. Cell Biol. 3, 317–27. doi:10.1038/nrm805
D'Arcy, A., Villard, F., Marsh, M., 2007. An automated microseed matrix-screening method for protein crystallization. Acta Crystallogr. Sect. D Biol. Crystallogr. 63, 550–554. doi:10.1107/S0907444907007652
Dahm, K., 2008. Role and regulation of human XRCC4-like factor/Cernunnos. J. Cell. Biochem. 104, 1534–1540. doi:10.1002/jcb.21726
Davies, A.A., Masson, J., Mcilwraith, M.J., Stasiak, A.Z., Stasiak, A., Venkitaraman, A.R., West, S.C., 2001. Role of BRCA2 in Control of the RAD51 Recombination and DNA Repair Protein 7, 273–282. doi:10.1016/S1097-2765(01)00175-7
Davies, O.R., Forment, J. V, Sun, M., Belotserkovskaya, R., Coates, J., Galanty, Y., Demir, M., Morton, C.R., Rzechorzek, N.J., Jackson, S.P., Pellegrini, L., 2015. CtIP tetramer assembly is required for DNA-end resection and repair. Nat. Struct. Mol. Biol. doi:10.1038/nsmb.2937
Davis, A.J., Chen, D.J., 2013. DNA double strand break repair via non-homologous end-joining. Transl. Cancer Res. 2, 130–143. doi:10.3978/j.issn.2218-676X.2013.04.02
Davis, I.W., Leaver-Fay, A., Chen, V.B., Block, J.N., Kapral, G.J., Wang, X., Murray, L.W., Arendall, W.B., Snoeyink, J., Richardson, J.S., Richardson, D.C., 2007. MolProbity: All-atom contacts and structure validation for proteins and nucleic acids. Nucleic Acids Res. 35, 375–383. doi:10.1093/nar/gkm216
de Boer, J., Hoeijmakers, J.H., 2000. Nucleotide excision repair and human syndromes. Carcinogenesis 21, 453–460. doi:10.1093/carcin/21.3.453
De Jager, M., Dronkert, M.L.G., Modesti, M., Beerens, C.E.M.T., Kanaar, R., Van Gent, D.C., 2001a. DNA-binding and strand-annealing activities of human Mre11: implications for its roles in DNA double-strand break repair pathways. Nucleic Acids Res. 29, 1317–1325. doi:10.1177/0194599812443221
De Jager, M., Van Noort, J., Van Gent, D.C., Dekker, C., Kanaar, R., Wyman, C., 2001b. Human Rad50/Mre11 is a flexible complex that can tether DNA ends. Mol. Cell 8, 1129–1135. doi:10.1016/S1097-2765(01)00381-1
de Klein, A., Muijtjens, M., van Os, R., Verhoeven, Y., Smit, B., Carr, A.M., Lehmann, A.R., Hoeijmakers, J.H.J., 2000. Targeted disruption of the cell-cycle checkpoint gene ATR leads to early embryonic lethality in mice. Curr. Biol. 10, 479–482. doi:10.1016/S0960-9822(00)00447-4
De Zio, D., Cianfanelli, V., Cecconi, F., 2013. New Insights into the Link Between DNA Damage and Apoptosis. Antioxid. Redox Signal. 19, 559–571. doi:10.1089/ars.2012.4938
Deans, A.J., West, S.C., 2011. DNA interstrand crosslink repair and cancer. Nat. Rev. Cancer 11, 467–480. doi:10.1038/nrc3088
Delacroix, S., Wagner, J.M., Kobayashi, M., Yamamoto, K.I., Karnitz, L.M., 2007. The Rad9-Hus1-Rad1 (9-1-1) clamp activates checkpoint signaling via TopBP1. Genes Dev. 21, 1472–1477. doi:10.1101/gad.1547007
Delorenzi, M., Speed, T., 2002. An HMM model for coiled-coil domains and a comparison with PSSM-based predictions. Bioinformatics 18, 617–625. doi:10.1093/bioinformatics/18.4.617
Derewenda, Z.S., Vekilov, P.G., 2006. Entropy and surface engineering in protein crystallization. Acta Crystallogr. Sect. D Biol. Crystallogr. 62, 116–124. doi:10.1107/S0907444905035237
Deshpande, R. a., Lee, J.H., Paull, T.T., 2017. Rad50 ATPase activity is regulated by DNA ends and requires coordination of both active sites. Nucleic Acids Res. 45, 5255–5268. doi:10.1093/nar/gkx173
Deshpande, R. a., Williams, G.J., Limbo, O., Williams, R.S., Kuhnlein, J., Lee, J.H., Classen, S., Guenther, G., Russell, P., Tainer, J. a., Paull, T.T., 2014. ATP-driven Rad50 conformations regulate DNA tethering, end resection, and ATM checkpoint signaling. EMBO J. 33, 482–500. doi:10.1002/embj.201386100

Dohmen, A.J.C., Qiao, X., Duursma, A., Wijdeven, R.H., Lieftink, C., Hageman, F., Morris, B., Halonen, P., Vens, C., van den Brekel, M.W.M., Ovaa, H., Neefjes, J., Zuur, C.L., 2017. Identification of a novel ATM inhibitor with cancer cell specific radiosensitization activity. Oncotarget 8, 73925–73937. doi:10.18632/oncotarget.18034

Dubin, M.J., Stokes, P.H., Sum, E.Y.M., Williams, R.S., Valova, V. a., Robinson, P.J., Lindeman, G.J., Glover, J.N.M., Visvader, J.E., Matthews, J.M., 2004. Dimerization of CtIP, a BRCA1- and CtBP-interacting protein, is mediated by an N-terminal coiled-coil motif. J. Biol. Chem. 279, 26932–26938. doi:10.1074/jbc.M313974200

Dulbecco, R., 1949. Reactivation of Ultra-Violet-Inactivated Bacteriophage by Visible Light. Nature 163, 949.

Duquette, M.L., Zhu, Q., Taylor, E.R., Tsay, A.J., Shi, L.Z., Berns, M.W., Mcgowan, C.H., 2012. CtIP Is Required to Initiate Replication-Dependent Interstrand Crosslink Repair 8, 26–28. doi:10.1371/journal.pgen.1003050

Durante, M., Loeffler, J.S., 2010. Charged particles in radiation oncology. Nat. Rev. Clin. Oncol. 7, 37–43. doi:10.1038/nrclinonc.2009.183

Duursma, A.M., Driscoll, R., Elias, J.E., Cimprich, K. a., 2013. A Role for the MRN Complex in ATR Activation via TOPBP1 Recruitment. Mol. Cell 50, 116–122. doi:10.1016/j.molcel.2013.03.006

Edlinger, B., Schlögelhofer, P., 2011. Have a break: Determinants of meiotic DNA double strand break (DSB) formation and processing in plants. J. Exp. Bot. 62, 1545–1563. doi:10.1093/jxb/erq421

Elmore, S., 2007. Apoptosis: a review of programmed cell death. Toxicol. Pathol. 35, 495–516. doi:10.1080/01926230701320337

Emsley, P., Cowtan, K., 2004. Coot: Model-building tools for molecular graphics. Acta Crystallogr. Sect. D Biol. Crystallogr. 60, 2126–2132. doi:10.1107/S0907444904019158

Errico, A., Costanzo, V., 2010. Differences in the DNA replication of unicellular eukaryotes and metazoans: Known unknowns. EMBO Rep. 11, 270–278. doi:10.1038/embor.2010.27

Evans, P.R., Murshudov, G.N., 2013. How good are my data and what is the resolution? Acta Crystallogr. Sect. D Biol. Crystallogr. 69, 1204–1214. doi:10.1107/S0907444913000061

Fan, C.H., Liu, W.L., Cao, H., Wen, C., Chen, L., Jiang, G., 2013. O6-methylguanine DNA methyltransferase as a promising target for the treatment of temozolomide-resistant gliomas. Cell Death Dis. 4, e876-8. doi:10.1038/cddis.2013.388

Fedier, A., Schwarz, V. a., Walt, H., Carpini, R.D., Haller, U., Fink, D., 2001. Resistance to topoisomerase poisons due to loss of DNA mismatch repair. Int. J. Cancer 93, 571–576. doi:10.1002/ijc.1356

Feigin, Svergun, D.I., 1989. Structure Analysis by Small-Angle X-Ray and Neutron Scattering. Acta Polym. doi:10.1002/actp.1989.010400317

Fell, V.L., Schild-Poulter, C., 2015. The Ku heterodimer: Function in DNA repair and beyond. Mutat. Res. - Rev. Mutat. Res. 763, 15–29. doi:10.1016/j.mrrev.2014.06.002

Feng, L., Chen, J., 2012. The E3 ligase RNF8 regulates KU80 removal and NHEJ repair. Nat. Struct. Mol. Biol. 19, 201–206. doi:10.1038/nsmb.2211

Ferretti, L.P., Himmels, S.-F., Trenner, A., Walker, C., von Aesch, C., Eggenschwiler, A., Murina, O., Enchev, R.I., Peter, M., Freire, R., Porro, A., Sartori, A. a., 2016. Cullin3-KLHL15 ubiquitin ligase mediates CtIP protein turnover to fine-tune DNA-end resection. Nat. Commun. 7, 12628. doi:10.1038/ncomms12628

Ferretti, L.P., Lafranchi, L., Sartori, A. a., 2013. Controlling DNA-end resection: A new task for CDKs. Front. Genet. 4, 1–7. doi:10.3389/fgene.2013.00099

Finn, R.S., Crown, J.P., Lang, I., Boer, K., Bondarenko, I.M., Kulyk, S.O., Ettl, J., Patel, R., Pinter, T., Schmidt, M., Shparyk, Y., Thummala, A.R., Voytko, N.L., Fowst, C., Huang, X., Kim, S.T., Randolph, S., Slamon, D.J., 2015. The cyclin-dependent kinase 4/6 inhibitor palbociclib in combination with letrozole versus letrozole alone as first-line treatment of oestrogen receptor-positive, HER2-negative, advanced breast cancer (PALOMA-1/TRIO-18): a randomised phase 2 study. Lancet Oncol. 16, 25–35. doi:10.1016/S1470-2045(14)71159-3

Fischer, A., Schumacher, N., Maier, M., Sendtner, M., Gessler, M., 2004. The Notch target genes Hey1 and Hey2 are required for embryonic vascular development. Genes Dev. 18, 901–911. doi:10.1101/gad.291004

Fokas, E., Prevo, R., Hammond, E.M., Brunner, T.B., McKenna, W.G., Muschel, R.J., 2014. Targeting ATR in DNA damage response and cancer therapeutics. Cancer Treat. Rev. 40, 109–117. doi:10.1016/j.ctrv.2013.03.002

Foray, N., Marot, D., Gabriel, A., Randrianarison, V., Carr, A.M., Perricaudet, M., Ashworth, A., Jeggo, P., 2003. A subset of ATM- and ATR-dependent phosphorylation events requires the BRCA1 protein. EMBO J. 22, 2860–2871. doi:10.1093/emboj/cdg274

Foster, S.S., Balestrini, a., Petrini, J.H.J., 2011. Functional Interplay of the Mre11 Nuclease and Ku in the Response to Replication-Associated DNA Damage. Mol. Cell. Biol. 31, 4379–4389. doi:10.1128/MCB.05854-11

Friedberg, E.C., 2008. A brief history of the DNA repair field. Cell Res. 18, 3–7. doi:10.1038/cr.2007.113

Friedberg, E.C., Walker, G.., Siede, W., Wood, R.D., Schultz, R.., Ellenberger, T., 2005. DNA Repair and Mutagenesis.

Fromme, J.C., Verdine, G.L., 2004. Base Excision Repair. Adv. Protein Chem. 69, 1–41. doi:10.1016/S0065-3233(04)69001-2

Fulda, S., Debatin, K.-M., 2006. Extrinsic versus intrinsic apoptosis pathways in anticancer chemotherapy. Oncogene 25, 4798.

Fusco, C., Reymond, A., Zervos, A.S., 1998. Molecular cloning and characterization of a novel Retinoblastoma-Binding Protein. Genomics 51, 351–358.

Ganai, R. a., Johansson, E., 2016. DNA Replication-A Matter of Fidelity. Mol. Cell 62, 745–755. doi:10.1016/j.molcel.2016.05.003

Ganem, N.J., Pellman, D., 2012. Linking abnormal mitosis to the acquisition of DNA damage. J. Cell Biol. 199, 871–881. doi:10.1083/jcb.201210040

Garcia, V., Phelps, S.E.L., Gray, S., Neale, M.J., 2011. Bidirectional resection of DNA double-strand breaks by Mre11 and Exo1. Nature 479, 241–4. doi:10.1038/nature10515

Germani, A., Prabel, A., Mourah, S., Podgorniak, M.P., Di Carlo, A., Ehrlich, R., Gisselbrecht, S., Varin-Blank, N., Calvo, F., Bruzzoni-Giovanelli, H., 2003. SIAH-1 interacts with CtIP and promotes its degradation by the proteasome pathway. Oncogene 22, 8845–8851. doi:10.1038/sj.onc.1206994

Giunta, S., Belotserkovskaya, R., Jackson, S.P., 2010. DNA damage signaling in response to double-strand breaks during mitosis. J. Cell Biol. 190, 197–207. doi:10.1083/jcb.200911156

Goldschmidt, L., Cooper, D.R., Derewenda, Z.S., Eisenberg, D., 2007. Toward rational protein crystallization: A Web server for the design of crystallizable protein variants. Protein Sci. 16, 1569–1576. doi:10.1110/ps.072914007

Gómez-Herreros, F., Romero-Granados, R., Zeng, Z., Álvarez-Quilón, A., Quintero, C., Ju, L., Umans, L., Vermeire, L., Huylebroeck, D., Caldecott, K.W., Cortés-Ledesma, F., 2013. TDP2-Dependent Non-Homologous End-Joining Protects against Topoisomerase II-Induced DNA Breaks and Genome Instability in Cells and In Vivo. PLoS Genet. 9. doi:10.1371/journal.pgen.1003226

Goodman, M.F., Woodgate, R., 2013. Translesion DNA Polymerase. Cold Spring Harb Perspect Biol 4, 247–250. doi:10.1101/cshperspect.a010363

Grabarz, A., Guirouilh-Barbat, J., Barascu, A., Pennarun, G., Genet, D., Rass, E., Germann, S.M., Bertrand, P., Hickson, I.D., Lopez, B.S., 2013. A Role for BLM in Double-Strand Break Repair Pathway Choice: Prevention of CtIP/Mre11-Mediated Alternative Nonhomologous End-Joining. Cell Rep. 5, 21–28. doi:10.1016/j.celrep.2013.08.034

Graham, A.I., Hunt, S., Stokes, S.L., Bramall, N., Bunch, J., Cox, A.G., McLeod, C.W., Poole, R.K., 2009. Severe zinc depletion of Escherichia coli: roles for high affinity zinc binding by ZinT, zinc transport and zinc-independent proteins. J. Biol. Chem. 284, 18377–89. doi:10.1074/jbc.M109.001503

Gravel, S., Chapman, J., Magill, C., Jackson, S., 2008. DNA helicases Sgs1 and BLM promote DNA double-strand break resection. Genes Dev. 22, 2767–2772. doi:10.1101/gad.503108
Graves, P.R., Lovly, C.M., Uy, G.L., Piwnica-Worms, H., 2001. Localization of human Cdc25C is regulated both by nuclear export and 14-3-3 protein binding. Oncogene 20, 1839.
Greenberg, R. a., Sobhian, B., Pathania, S., Cantor, S.B., Nakatani, Y., Livingston, D.M., 2006. Multifactorial contributions to an acute DNA damage response by BRCA1/BARD1-containing complexes. Genes Dev. 20, 34–46. doi:10.1101/gad.1381306
Gu, B., Chen, P.-L., 2014. Expression of PCNA-binding domain of CtIP, a motif required for CtIP localization at DNA replication foci, causes DNA damage and activation of DNA damage checkpoint. Cell Cycle 8, 1409–1420. doi:10.4161/cc.8.9.8322
Hada, M., Georgakilas, A.G., 2008. Formation of Clustered DNA Damage after High-LET Irradiation: A Review. J. Radiat. Res. 49, 203–210. doi:10.1269/jrr.07123
Hampton Research, 2001. Temperature as a Crystallization Variable Temperature as a Crystallization Variable. Solut. Cryst. Growth 2.
Hanawalt, P.C., 1994. Transcription-Coupled Repair and human disease 266.
Hao, J., de Renty, C., Li, Y., Xiao, H., Kemp, M.G., Han, Z., DePamphilis, M.L., Zhu, W., 2015. And-1 coordinates with Claspin for efficient Chk1 activation in response to replication stress. EMBO J. 34, 1845–1985. doi:10.15252/embj.201488016
Harding, S.E., Jumel, K., 1998. Light Scattering. Curr. Protoc. Protein Sci. 11, 7.8.1-7.8.14. doi:10.1002/0471140864.ps0708s11
Harper, J.W., Elledge, S.J., 2007. The DNA Damage Response: Ten Years After. Mol. Cell 28, 739–745. doi:10.1016/j.molcel.2007.11.015
Harper, J.W., Elledge, S.J., Keyomarsi, K., Dynlacht, B., Tsai, L.H., Zhang, P., Dobrowolski, S., Bai, C., Connell-Crowley, L., Swindell, E., 1995. Inhibition of cyclin-dependent kinases by p21. Mol. Biol. Cell 6, 387–400. doi:10.1111/j.1365-2184.2009.00584.x
Hartlerode, A.J., Morgan, M.J., Wu, Y., Buis, J., Ferguson, D.O., 2016. Recruitment and activation of the ATM kinase in the absence of DNA damage sensors. Nat struct mol biol 8, 583–592. doi:10.1002/aur.1474.Replication
Hartsuiker, E., Neale, M.J., Carr, A.M., 2009. Distinct requirements for the Rad32(Mre11) nuclease and Ctp1(CtIP) in the removal of covalently bound topoisomerase I and II from DNA. 33, 117–123. doi:10.1016/j.molcel.2008.11.021.Distinct
Hecht, stephen S., 1999. Tobacco Smoke Carcinogens and Lung Cancer.[Review]. J. Natl. Cancer Inst. 91, 1194–1210.
Hefferin, M.L., Tomkinson, A.E., 2005. Mechanism of DNA double-strand break repair by non-homologous end joining. DNA Repair (Amst). 4, 639–648. doi:10.1016/j.dnarep.2004.12.005
Her, J., Bunting, S.F., 2018. How cells ensure correct repair of DNA double-strand breaks. J. Biol. Chem. 293, 10502–10511. doi:10.1074/jbc.TM118.000371
Heyer, W.-D., Ehmsen, K.T., Liu, J., 2010. Regulation of Homologous Recombination in Eukaryotes. Annu. Rev. Genet. 44, 113–139. doi:10.1146/annurev-genet-051710-150955
Hirai, H., Iwasawa, Y., Okada, M., Arai, T., Nishibata, T., Kobayashi, M., Kimura, T., Kaneko, N., Ohtani, J., Yamanaka, K., Itadani, H., Takahashi-Suzuki, I., Fukasawa, K., Oki, H., Nambu, T., Jiang, J., Sakai, T., Arakawa, H., Sakamoto, T., Sagara, T., Yoshizumi, T., Mizuarai, S., Kotani, H., 2009. Small-molecule inhibition of Wee1 kinase by MK-1775 selectively sensitizes p53-deficient tumor cells to DNA-damaging agents. Mol. Cancer Ther. 8, 2992–3000. doi:10.1158/1535-7163.MCT-09-0463
Hoeijmakers, J.H.J., 2009. DNA damage, aging, and cancer. N. Engl. J. Med. 361, 1475–85. doi:10.1056/NEJMra0804615
Hohl, M., Kwon, Y., Galván, S.M., Xue, X., Tous, C., Aguilera, A., Sung, P., Petrini, J.H.J., 2010. The Rad50 coiled-coil domain is indispensable for Mre11 complex functions. Nat. Struct. Mol. Biol. 18, 1124–1131. doi:10.1038/nsmb.2116
Holliday, R., 1964. A mechanism for gene conversion in fungi. Genet. Res. 5, 282–304. doi:DOI:

10.1017/S0016672300001233
Holzer, S., 2017. Priming and Pausing DNA Replication.
Hopfner, K.-P., Craig, L., Moncalian, G., Zinkel, R.A., Usui, T., Owen, B.A.L., Karcher, A., Henderson, B., Bodmer, J.-L., McMurray, C.T., Carney, J.P., Petrini, J.H.J., Tainer, J.A., 2002. The Rad50 zinc-hook is a structure joining Mre11 complexes in DNA recombination and repair. Nature 418, 562.
Hsieh, P., Zhang, Y., 2017. The Devil is in the details for DNA mismatch repair. Proc. Natl. Acad. Sci. 114, 3552–3554. doi:10.1073/pnas.1702747114
Huen, M.S.Y., Sy, shirley M.., Chen, J., 2010. BRCA1 and its toolbox for the maintenance of genome integrity. Nat. Rev. Mol. Cell Biol. 11, 138–148. doi:10.1038/nrm2831
Huertas, P., Cortés-Ledesma, F., Sartori, A. a., Aguilera, A., Jackson, S.P., 2008. CDK targets Sae2 to control DNA-end resection and homologous recombination. Nature 455, 689–692. doi:10.1038/nature07215
Huertas, P., Jackson, S.P., 2009. Human CtIP mediates cell cycle control of DNA end resection and double strand break repair. J. Biol. Chem. 284, 9558–65. doi:10.1074/jbc.M808906200
Hunt, J.B., Neece, S.U.E.H., Ginsburg, a N.N., 1985. The Use of 4- (2-Pyridylazo) resorcinol in Studies of Zinc Release from Escherichia co / i Aspartate Transcarbamoylase ’. Anal. Biochem. 157, 150–157.
Huynh, K., Partch, C.L., 2015. Analysis of protein stability and ligand interactions by thermal shift assay. Curr. Protoc. protein Sci. 79, 28.9.1-28.9.14. doi:10.1002/0471140864.ps2809s79
Ishiai, M., Kitao, H., Smogorzewska, A., Tomida, J., Kinomura, A., Uchida, E., Saberi, A., Kinoshita, E., Kinoshita-Kikuta, E., Koike, T., Tashiro, S., Elledge, S.J., Takata, M., 2008. FANCI phosphorylation functions as a molecular switch to turn on the Fanconi anemia pathway. Nat. Struct. Mol. Biol. 15, 1138–1146. doi:10.1038/nsmb.1504
Jackson, S.P., Bartek, J., 2009. The DNA-damage response in human biology and disease. Nature 461, 1071–1078. doi:10.1038/nature08467
Jasin, M., Rothstein, R., 2013. Repair of strand breaks by homologous recombination. Cold Spring Harb. Perspect. Biol. 5, 1–18. doi:10.1101/cshperspect.a012740
Javle, M., Curtin, N.J., 2011. The role of PARP in DNA repair and its therapeutic exploitation. Br. J. Cancer 105, 1114–1122. doi:10.1038/bjc.2011.382
Jazayeri, A., Falck, J., Lukas, C., Bartek, J., Smith, G.C.M., Lukas, J., Jackson, S.P., 2006. ATM- and cell cycle-dependent regulation of ATR in response to DNA double-strand breaks. Nat. Cell Biol. 8, 37–45. doi:10.1038/ncb1337
Jeon, Y., Ko, E., Lee, K.Y., Ko, M.J., Park, S.Y., Kang, J., Jeon, C.H., Lee, H., Hwang, D.S., 2011. TopBP1 deficiency causes an early embryonic lethality and induces cellular senescence in primary cells. J. Biol. Chem. 286, 5414–5422. doi:10.1074/jbc.M110.189704
Jezorek, J.R., Freiser, H., 1979. 4-(Pyridylazo)resorcinol-Based Continuous Detection System for Trace Levels of Metal Ions. Anal. Chem. 51, 373–376. doi:10.1021/ac50039a012
Jimeno, S., Fernandez-Avila, M.J., Cruz-Garcia, a., Cepeda-Garcia, C., Gomez-Cabello, D., Huertas, P., 2015. Neddylation inhibits CtIP-mediated resection and regulates DNA double strand break repair pathway choice. Nucleic Acids Res. 1–13. doi:10.1093/nar/gku1384
Jin, S., Antinore, M.J., Lung, F.D.T., Dong, X., Zhao, H., Fan, F., Colchagie, A.B., Blanck, P., Roller, P.P., Fornace, A.J., Zhan, Q., 2000. The GADD45 inhibition of Cdc2 kinase correlates with GADD45-mediated growth suppression. J. Biol. Chem. 275, 16602–16608. doi:10.1074/jbc.M000284200
John, P.C.L., Mews, M., Moore, R., 2001. Cyclin/Cdk complexes: Their involvement in cell cycle progression and mitotic division. Protoplasma 216, 119–142. doi:10.1007/BF02673865
Kabsch, W., 2010. *XDS*. Acta Crystallogr. Sect. D Biol. Crystallogr. 66, 125–132. doi:10.1107/S0907444909047337

Kaidi, A., Jackson, S.P., 2013. KAT5 tyrosine phosphorylation couples chromatin sensing to ATM signalling. Nature 498, 70–74. doi:10.1038/nature12201
Kaidi, A., Weinert, B.T., Choudhary, C., Jackson, S.P., 2010. Human SIRT6 promotes DNA end resection through CtIP deacetylation. Science 329, 1348–53. doi:10.1126/science.1192049

The above two papers have now been retracted

Kaniecki, K., De Tullio, L., Greene, E.C., 2018. A change of view: Homologous recombination at single-molecule resolution. Nat. Rev. Genet. 19, 191–207. doi:10.1038/nrg.2017.92
Karmakar, P., Snowden, C.M., Ramsden, D. a., Bohr, V. a., 2002. Ku heterodimer binds to both ends of the Werner protein and functional interaction occurs at the Werner N-terminus. Nucleic Acids Res. 30, 3583–91. doi:doi: 10.1093/nar/gkf482
Kelner, a, 1949. Effect of visible light on the recovery of Streptomyces griseus conidia from ultraviolet irradiation injury. Proc. Natl. Acad. Sci. U. S. A. 35, 73–79. doi:10.1073/pnas.35.2.73
Kijas, A.W., Lim, Y.C., Bolderson, E., Cerosaletti, K., Gatei, M., Jakob, B., Tobias, F., Taucher-Scholz, G., Gueven, N., Oakley, G., Concannon, P., Wolvetang, E., Khanna, K.K., Wiesm??ller, L., Lavin, M.F., 2015. ATM-dependent phosphorylation of MRE11 controls extent of resection during homology directed repair by signalling through Exonuclease 1. Nucleic Acids Res. 43, 8352–8367. doi:10.1093/nar/gkv754
Kilkenny, M.L., Simon, A.C., Mainwaring, J., Wirthensohn, D., Holzer, S., Pellegrini, L., 2017. The human CTF4-orthologue AND-1 interacts with DNA polymerase a/primase via its unique C-Terminal HMG box. J. R. Soc. Interface 14. doi:10.1098/rsob.170217
Kim, B.W., Jung, Y.O., Kim, M.K., Kwon, D.H., Park, S.H., Kim, J.H., Kuk, Y.B., Oh, S.J., Kim, L., Kim, B.H., Yang, W.S., Song, H.K., 2017. ACCORD: An assessment tool to determine the orientation of homodimeric coiled-coils. Sci. Rep. 7, 1–11. doi:10.1038/srep43318
Kim, H.S., Vijayakumar, S., Reger, M., Harrison, J.C., Haber, J.E., Weil, C., Petrini, J.H.J., 2008. Functional interactions between Sae2 and the Mre11 complex. Genetics 178, 711–723. doi:10.1534/genetics.107.081331
Kim, Y.-J., M. Wilson III, D., 2012. Overview of Base Excision Repair Biochemistry. Curr. Mol. Pharmacol. 5, 3–13. doi:10.2174/1874467211205010003
Kim, Y., Starostina, N.G., Kipreos, E.T., 2008. The CRL4Cdt2 ubiquitin ligase targets the degradation of p21Cip1 to control replication licensing. Genes Dev. 22, 2507–2519. doi:10.1101/gad.1703708
Kitadai, N., Maruyama, S., 2018. Origins of building blocks of life: A review. Geosci. Front. 9, 1117–1153. doi:10.1016/j.gsf.2017.07.007
Knauer, S.K., Mahendrarajah, N., Roos, W.P., Krämer, O.H., 2015. The inducible E3 ubiquitin ligases SIAH1 and SIAH2 perform critical roles in breast and prostate cancers. Cytokine Growth Factor Rev. 26, 405–413. doi:10.1016/j.cytogfr.2015.04.002
Knoch, J., Kamenisch, Y., Kubisch, C., Bemeburg, M., 2012. Rare hereditary diseases with defects in DNA-repair. Eur. J. Dermatology 22, 443–455.
Kobayashi, J., Okui, M., Asaithamby, A., Burma, S., Chen, B.P.C., Tanimoto, K., Matsuura, S., Komatsu, K., Chen, D.J., 2010. WRN participates in translesion synthesis pathway through interaction with NBS1. Mech. Ageing Dev. 131, 436–444. doi:10.1016/j.mad.2010.06.005
Köcher, S., Spies-Naumann, A., Kriegs, M., Dahm-Daphi, J., Dornreiter, I., 2013. ATM is required for the repair of Topotecan-induced replication-associated double-strand breaks. Radiother. Oncol. 108, 409–414. doi:10.1016/j.radonc.2013.06.024
Koipally, J., Georgopoulos, K., 2002. Ikaros-CtIP interactions do not require C-terminal binding protein and participate in a deacetylase-independent mode of repression. J. Biol. Chem. 277, 23143–23149. doi:10.1074/jbc.M202079200
Kowalczykowski, S.C., 2015. An Overview of the Molecular Mechanisms of Recombinational DNA Repair. Cold Spring Harb. Perspect. Biol. 7, a016410. doi:10.1101/cshperspect.a016410
Krejci, L., Altmannova, V., Spirek, M., Zhao, X., 2012. Homologous recombination and its regulation.

Nucleic Acids Res. 40, 5795–818. doi:10.1093/nar/gks270
Krokan and Bjoras, 2013. Base Excision Repair 1–17.
Kumagai, A., Kim, S.M., Dunphy, W.G., 2004. Claspin and the activated form of ATR-ATRIP collaborate in the activation of Chk1. J. Biol. Chem. 279, 49599–49608. doi:10.1074/jbc.M408353200
Kumagai, A., Lee, J., Yoo, H.Y., Dunphy, W.G., 2006. TopBP1 activates the ATR-ATRIP complex. Cell 124, 943–955. doi:10.1016/j.cell.2005.12.041
Lafarga, V., Cuadrado, a., Lopez de Silanes, I., Bengoechea, R., Fernandez-Capetillo, O., Nebreda, a. R., 2009. p38 Mitogen-Activated Protein Kinase- and HuR-Dependent Stabilization of p21Cip1 mRNA Mediates the G1/S Checkpoint. Mol. Cell. Biol. 29, 4341–4351. doi:10.1128/MCB.00210-09
Lafarga, V., Cuadrado, A., Lopez de Silanes, I., Bengoechea, R., Fernandez-Capetillo, O., Nebreda, A.R., 2009. p38 Mitogen-Activated Protein Kinase- and HuR-Dependent Stabilization of p21Cip1 mRNA Mediates the G1/S Checkpoint. Mol. Cell. Biol. 29, 4341 LP-4351.
Lafrance-Vanasse, J., Williams, G.J., Tainer, J. a., 2015. Envisioning the dynamics and flexibility of Mre11-Rad50-Nbs1 complex to decipher its roles in DNA replication and repair. Prog. Biophys. Mol. Biol. 117, 182–193. doi:10.1016/j.pbiomolbio.2014.12.004
Lafranchi, L., Boer, H.R. De, Vries, E.G.E. De, Ong, S., Sartori, A.A., Vugt, M.A.T.M. Van, 2014. APC / C Cdh 1 controls CtIP stability during the cell cycle and in response to DNA damage 1–20.
Lamarche, B.J., Orazio, N.I., Weitzman, M.D., 2010. The MRN complex in double-strand break repair and telomere maintenance. FEBS Lett. 584, 3682–3695. doi:10.1016/j.febslet.2010.07.029
Lammens, K., Bemeleit, D.J., Möckel, C., Clausing, E., Schele, A., Hartung, S., Schiller, C.B., Lucas, M., Angermüller, C., Söding, J., Sträßer, K., Hopfner, K.P., 2011. The Mre11:Rad50 structure shows an ATP-dependent molecular clamp in DNA double-strand break repair. Cell 145, 54–66. doi:10.1016/j.cell.2011.02.038
Larkin, M. a., Blackshields, G., Brown, N.P., Chenna, R., Mcgettigan, P. a., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, a., Lopez, R., Thompson, J.D., Gibson, T.J., Higgins, D.G., 2007. Clustal W and Clustal X version 2.0. Bioinformatics 23, 2947–2948. doi:10.1093/bioinformatics/btm404
Lavin, M., Kozlov, S., Gatei, M., Kijas, A., 2015. ATM-Dependent Phosphorylation of All Three Members of the MRN Complex: From Sensor to Adaptor. Biomolecules 5, 2877–2902. doi:10.3390/biom5042877
Lee, J.-H., Paull, T.T., 2004. Direct Activation of the ATM Protein Kinase by the Mre11/Rad50/Nbs1 Complex. Science (80-.). 304, 93 LP-96.
Lee, J.H., Mand, M.R., Deshpande, R. a., Kinoshita, E., Yang, S.H., Wyman, C., Paull, T.T., 2013. Ataxia Telangiectasia-Mutated (ATM) kinase activity is regulated by ATP-driven conformational changes in the Mre11/Rad50/Nbs1 (MRN) complex. J. Biol. Chem. 288, 12840–12851. doi:10.1074/jbc.M113.460378
Lee, J.H., Paull, T.T., 2006. Purification and Biochemical Characterization of Ataxia-Telangiectasia Mutated and Mre11/Rad50/Nbs1. Methods Enzymol. 408, 529–539. doi:10.1016/S0076-6879(06)08033-5
Lee, J.H., Paull, T.T., 2005. ATM activation by DNA double-strand breaks through the Mre11-Rad50-Nbs1 complex. Science (80-.). 308, 551–554. doi:10.1126/science.1108297
Lee, W., Watters, K.E., Troupis, A.T., Reinen, N.M., Suchy, F.P., Moyer, K.L., Frederick, R.O., Tonelli, M., Aceti, D.J., Palmenberg, A.C., Markley, J.L., 2014. Solution structure of the 2A protease from a common cold agent, human rhinovirus C2, strain W12. PLoS One 9, e97198. doi:10.1371/journal.pone.0097198
Lengsfeld, B.M., J., R.A., Bhaskara, V., Ghirlando, R., Paull, T.T., 2007. Sae2 is an endonuclease that processes hairpin DNA cooperatively with the Mre11/Rad50/Xrs2 complex 33, 1212–1217. doi:10.1016/j.dci.2009.07.003.Characterization
Li, G.-M., 2007. Mechanisms and functions of DNA mismatch repair. Cell Res. 18, 85.
Li, S., Chen, P.L., Subramanian, T., Chinnadurai, G., Tomlinson, G., Osborne, C.K., Sharp, Z.D., Lee,

W.H., 1999. Binding of CtIP to the BRCT repeats of BRCA1 involved in the transcription regulation of p21 is disrupted upon DNA damage. J. Biol. Chem. 274, 11334–11338. doi:10.1074/jbc.274.16.11334
Li, S., Ting, N.S.Y., Zheng, L., Chen, P., Ziv, Y., Shiloh, Y., Lee, E.Y.P., Lee, W., 2000. Functional link of BRCA1 and ataxia telangiectasia gene product in DNA damage response 406, 2–7.
Li, X., Heyer, W.D., 2008. Homologous recombination in DNA repair and DNA damage tolerance. Cell Res. 18, 99–113. doi:10.1038/cr.2008.1
Liao, S., Tammaro, M., Yan, H., 2016. The structure of ends determines the pathway choice and Mre11 nuclease dependency of DNA double-strand break repair. Nucleic Acids Res. 44, 5689–5701. doi:10.1093/nar/gkw274
Lim, H.S., Kim, J.S., Park, Y.B., Gwon, G.H., Cho, Y., 2011. Crystal structure of the Mre11-Rad50-ATPase complex: understanding the interplay between Mre11 and Rad50. Genes Dev. 25, 1091–1104. doi:10.1101/gad.2037811
Limbo, O., Chahwan, C., Yamada, Y., Bruin, R. a M. De, Russell, P., 2007. Ctp1 is a Cell Cycle-Regulated Protein that Functions with Mre11 Complex to Control Double-Strand Break Repair by Homologous Recombination. October 28, 134–146.
Liu, B., Cong, R., Peng, B., Zhu, B., Dou, G., Ai, H., Zhang, X., Wang, Z., Xu, X., 2014. CtIP is required for DNA damage-dependent induction of P21 1–6.
Liu, F., Lee, W.-H., 2006. CtIP activates its own and cyclin D1 promoters via the E2F/RB pathway during G1/S progression. Mol. Cell. Biol. 26, 3124–34. doi:10.1128/MCB.26.8.3124-3134.2006
Liu, F., Lee, W.-H., 2005. Phosphorylation of CtIP by ATM regulates its own expression and activates both G1/S and G2/M checkpoints. Cancer Res. 65, 606 LP-606.
Lloyd, J., Chapman, J.R., Clapperton, J. a, Haire, L.F., Hartsuiker, E., Li, J., Carr, A.M., Jackson, S.P., Smerdon, S.J., 2009. A supramodular FHA/BRCT-repeat architecture mediates Nbs1 adaptor function in response to DNA damage. Cell 139, 100–11. doi:10.1016/j.cell.2009.07.043
Lobley, C.M.C., Sandy, J., Sanchez-Weatherby, J., Mazzorana, M., Krojer, T., Nowak, R.P., Sorensen, T.L., 2016. A generic protocol for protein crystal dehydration using the HC1b humidity controller. Acta Crystallogr. Sect. D Struct. Biol. 72, 629–640. doi:10.1107/S2059798316003065
Lopez-Girona, A., Furnari, B., Mondesert, O., Russell, P., 1999. Nuclear localization of Cdc25 is regulated by DNA damage and a 14-3-3 protein. Nature 397, 172.
Lundblad, J.R., Laurance, M., Goodman, R.H., 2008. Fluorescence of Protein-DNA Interactions Polarization Analysis and Protein-Protein. Library (Lond). 607–612.
Lupas, A., Van Dyke, M., Stock, J., 1991. Predicting coiled coils from protein sequences. Science (80-.). 252, 1162 LP-1164.
Ma, Y., Lu, H., Tippin, B., Goodman, M.F., Shimazaki, N., Koiwai, O., Hsieh, C.L., Schwarz, K., Lieber, M.R., 2004. A biochemically defined system for mammalian nonhomologous DNA end joining. Mol. Cell 16, 701–713. doi:10.1016/j.molcel.2004.11.017
Ma, Y., Pannicke, U., Schwarz, K., Lieber, M.R., 2002. Hairpin opening and overhang processing by an Artemis/DNA-dependent protein kinase complex in nonhomologous end joining and V(D)J recombination. Cell 108, 781–794. doi:10.1016/S0092-8674(02)00671-2
Mächtle, W., Börger, L., 2006. Analytical Ultracentrifugation of Polymers and Nanoparticles. Springer-Verlag. doi:https://doi.org/10.1007/s00216-006-0491-5
Mahajan, K.N., McElhinny, S. a. N., Mitchell, B.S., Ramsden, D. a., 2002. Association of DNA polymerase mu (pol mu) with Ku and ligase IV: role for pol mu in end-joining double-strand break repair. Mol. Cell. Biol. 22, 5194–5202. doi:10.1128/MCB.22.14.5194
Makharashvili, N., Tubbs, A.T., Yang, S.-H., Wang, H., Barton, O., Zhou, Y., Deshpande, R. a, Lee, J.-H., Lobrich, M., Sleckman, B.P., Wu, X., Paull, T.T., 2014. Catalytic and noncatalytic roles of the CtIP endonuclease in double-strand break end resection. Mol. Cell 54, 1022–33. doi:10.1016/j.molcel.2014.04.011
Mao, Z., Bozzella, M., Seluanov, A., Gorbunova, V., 2008. Comparison of nonhomologous end joining and homologous recombination in human cells. DNA Repair (Amst). 7, 1765–1771.

doi:10.1016/j.dnarep.2008.06.018
Marini, V., Krejci, L., 2010. Srs2: The “Odd-Job Man” in DNA repair. DNA Repair (Amst). 9, 268–275. doi:10.1016/j.dnarep.2010.01.007
Marteijn, J.A., Lans, H., Vermeulen, W., Hoeijmakers, J.H.J., 2014. Understanding nucleotide excision repair and its roles in cancer and ageing. Nat. Rev. Mol. Cell Biol. 15, 465.
Mateos-Gomez, P. a., Gong, F., Nair, N., Miller, K.M., Lazzerini-Denchi, E., Sfeir, A., 2015. Mammalian polymerase θ promotes alternative NHEJ and suppresses recombination. Nature 518, 254–257. doi:10.1038/nature14157
Matsuoka, S., Huang, M., Elledge, S.J., 1998. Linkage of ATM to Cell Cycle Regulation by the Chk2 Protein Kinase. Science (80-.). 282, 1893 LP-1897.
McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., Read, R.J., 2007. Phaser crystallographic software. J. Appl. Crystallogr. 40, 658–674. doi:10.1107/S0021889807021206
McDonnell, a. V., Jiang, T., Keating, a. E., Berger, B., 2006. Paircoil2: Improved prediction of coiled coils from sequence. Bioinformatics 22, 356–358. doi:10.1093/bioinformatics/bti797
Mellon, I., Bohr, V. a, Smith, C. a, Hanawalt, P.C., 1986. Preferential {DNA} repair of an active gene in human cells. Proc. Natl. Acad. Sci. U. S. A. 83, 8878–8882. doi:10.1073/pnas.83.23.8878
Memisoglu, A., Samson, L., 2000. Base excision repair in yeast and mammals. Mutat. Res. - Fundam. Mol. Mech. Mutagen. 451, 39–51. doi:10.1016/S0027-5107(00)00039-7
Menzel, T., Näsignhse-Kumpf, V., Kousholt, A.N., Klein, D.K., Lund-Andersen, C., Lees, M., Johansen, J.V., Syljuåsen, R.G., Sørensen, C.S., 2011. A genetic screen identifies BRCA2 and PALB2 as key regulators of G2 checkpoint maintenance. EMBO Rep. 12, 705–712. doi:10.1038/embor.2011.99
Meselson, M.S., Radding, C.M., 1975. A general model for genetic recombination. Proc Natl Acad Sci U S A 72, 358–361. doi:10.1073/pnas.72.1.358
Meyerson, M., Enders, G.H., Wu, C.L., Su, L.K., Gorka, C., Nelson, C., Harlow, E., Tsai, L.H., 1992. A family of human cdc2-related protein kinases. EMBO J. 11, 2909–2917. doi:10.1002/j.1460-2075.1992.tb05360.x
Mimitou, E.P., Symington, L.S., 2009. DNA end resection: Many nucleases make light work. DNA Repair (Amst). 8, 983–995. doi:https://doi.org/10.1016/j.dnarep.2009.04.017
Moldovan, G.L., Dejsuphong, D., Petalcorin, M.I.R., Hofmann, K., Takeda, S., Boulton, S.J., D’Andrea, A.D., 2012. Inhibition of homologous recombination by the PCNA-interacting protein PARI. Mol. Cell 45, 75–86. doi:10.1016/j.molcel.2011.11.010
Moncalian, G., Lengsfeld, B., Bhaskara, V., Hopfner, K.P., Karcher, A., Alden, E., Tainer, J. a., Paull, T.T., 2004. The Rad50 Signature Motif: Essential to ATP Binding and Biological Function. J. Mol. Biol. 335, 937–951. doi:10.1016/j.jmb.2003.11.026
Moreau, S., Ferguson, J.R., Symington, L.S., 1999. The Nuclease Activity of Mre11 Is Required for Meiosis but Not for Mating Type Switching, End Joining, or Telomere Maintenance. Mol. Cell. Biol. 19, 556–566. doi:10.1128/MCB.19.1.556
Morgan, D.O., 1995. Principles of CDK regulation.pdf. doi:10.1038/374131a0
Morrical, S.W., 2016. DNA-Pairing and Annealing Processes HR HDR. Cold Spring Harb Perspect Biol 7, 1–20. doi:10.1101/cshperspect.a016444
Muniandy, P., Liu, J., Majumdar, A., Liu, S., Seidman, M.M., 2011. DNA Interstrand Crossling Repair in Mammalian Cells: Step by Step, Critical Reviews in Biochemistry and Molecular Biology. doi:10.3109/10409230903501819.DNA
Murina, O., von Aesch, C., Karakus, U., Ferretti, L.P., Bolck, H. a, Hänggi, K., Sartori, A. a, 2014. FANCD2 and CtIP cooperate to repair DNA interstrand crosslinks. Cell Rep. 7, 1030–8. doi:10.1016/j.celrep.2014.03.069
Murray, A.W., 2004. recycling the cell cyclins revisited 116, 14.
Myler, L.R., Gallardo, I.F., Soniat, M.M., Deshpande, R. a., Gonzalez, X.B., Kim, Y., Paull, T.T., Finkelstein, I.J., 2017. Single-Molecule Imaging Reveals How Mre11-Rad50-Nbs1 Initiates DNA Break Repair. Mol. Cell 67, 891–898.e4. doi:10.1016/j.molcel.2017.08.002

Nassif, N., Penney, J., Pal, S., Engels, W.R., Gloor, G.B., 1994. Efficient copying of nonhomologous sequences from ectopic sites via P-element-induced gap repair. Mol. Cell. Biol. 14, 1613–25. doi:10.1128/MCB.14.3.1613.

Norbury, C.J., Hickson, I.D., 2001. Cellular responses to dna damage 367–401.

Nyborg, J.K., Peersen, O.B., 2004. That zincing feeling: the effects of EDTA on the behaviour of zinc-binding transcriptional regulators. Biochem. J. 381, e3. doi:10.1042/BJ20041096

O'Driscoll, M., Jeggo, P.A., 2006. The role of double-strand break repair — insights from human genetics. Nat. Rev. Genet. 7, 45.

Ochi, T., Blackford, A.N., Coates, J., Jhujh, S., Mehmood, S., Tamura, N., Travers, J., Wu, Q., Draviam, V.M., Robinson, C. V, Tom, L., Jackson, S.P., 2015. PAXX , a paralog of XRCC4 and XLF , interacts with Ku to promote DNA double-strand break repair **. Science 347, 185–188. doi:10.1126/science.1261971.PAXX

Olson, E., Nievera, C.J., Lee, A.Y.L., Chen, L., Wu, X., 2007. The Mre11-Rad50-Nbs1 complex acts both upstream and downstream of ataxia telangiectasia mutated and Rad3-related protein (ATR) to regulate the S-phase checkpoint following UV treatment. J. Biol. Chem. 282, 22939–22952. doi:10.1074/jbc.M702162200

Orelli, B., McClendon, T.B., Tsodikov, O. V., Ellenberger, T., Niedernhofer, L.J., Schärer, O.D., 2010. The XPA-binding domain of ERCC1 is required for nucleotide excision repair but not other DNA repair pathways. J. Biol. Chem. 285, 3705–3712. doi:10.1074/jbc.M109.067538

Oswald, F., Winkler, M., Cao, Y., Astrahantseff, K., Bourteele, S., Borggrefe, T., Al, O.E.T., Iol, M.O.L.C.E.L.L.B., 2005. RBP-J / SHARP Recruits CtIP / CtBP Corepressors To Silence Notch Target Genes 25, 10379–10390. doi:10.1128/MCB.25.23.10379

Park, Y.B., Chae, J., Kim, Y.C., Cho, Y., 2011. Crystal structure of human Mre11: Understanding tumorigenic mutations. Structure 19, 1591–1602. doi:10.1016/j.str.2011.09.010

Parthasarathy, G., Cummings, R., Becker, J.W., Soisson, S.M., 2008. Surface-entropy reduction approaches to manipulate crystal forms of β-ketoacyl acyl carrier protein synthase II from Streptococcus pneumoniae. Acta Crystallogr. Sect. D Biol. Crystallogr. 64, 141–148. doi:10.1107/S090744490705559X

Parvathaneni, S., Stortchevoi, A., Sommers, J. a., Brosh, R.M., Sharma, S., 2013. Human RECQ1 Interacts with Ku70/80 and Modulates DNA End-Joining of Double-Strand Breaks. PLoS One 8. doi:10.1371/journal.pone.0062481

Paull, T.T., Deshpande, R.A., 2014. The Mre11/Rad50/Nbs1 complex: recent insights into catalytic activities and ATP-driven conformational changes. Exp. Cell Res. 329, 139–147. doi:10.1016/j.yexcr.2014.07.007

Paull, T.T., Gellert, M., 1998. The 3' to 5' exonuclease activity of Mre11 facilitates repair of DNA double-strand breaks. Mol. Cell 1, 969–979. doi:10.1016/S1097-2765(00)80097-0

Pernas, S., Tolaney, S.M., Winer, E.P., Goel, S., 2018. CDK4/6 inhibition in breast cancer: current practice and future directions. Ther. Adv. Med. Oncol. 10, 1758835918786451. doi:10.1177/1758835918786451

Peterson, S.E., Li, Y., Wu-Baer, F., Chait, B.T., Baer, R., Yan, H., Gottesman, M.E., Gautier, J., 2013. Activation of DSB processing requires phosphorylation of CtIP by ATR. Mol. Cell 49, 657–67. doi:10.1016/j.molcel.2012.11.020

Petoukhov, M. V, Franke, D., Shkumatov, A. V, Tria, G., Kikhney, A.G., Gajda, M., Gorba, C., Mertens, H.D.T., Konarev, P. V, Svergun, D.I., 2012. New developments in the ATSAS program package for small-angle scattering data analysis. J. Appl. Crystallogr. 45, 342–350. doi:10.1107/S0021889812007662

Petrini, J.H.J., Bressan, D.A., Yao, M.S., 1997. The RAD52 epistasis group in mammalian double strand break repair 9, 181–188.

Pluciennik, a., Dzantiev, L., Iyer, R.R., Constantin, N., Kadyrov, F. a., Modrich, P., 2010. PCNA function in the activation and strand direction of MutL endonuclease in mismatch repair. Proc. Natl. Acad. Sci. 107, 16066–16071. doi:10.1073/pnas.1010662107

Polato, F., Callen, E., Wong, N., Faryabi, R., Bunting, S., Chen, H.-T., Kozak, M., Kruhlak, M.J., Reczek, C.R., Lee, W.-H., Ludwig, T., Baer, R., Feigenbaum, L., Jackson, S., Nussenzweig, A., 2014. CtIP-mediated resection is essential for viability and can operate independently of BRCA1. J. Exp. Med. 211, 1027–36. doi:10.1084/jem.20131939

Polo, S., Jackson, S., 2011. Dynamics of DNA damage response proteins at DNA breaks: a focus on protein modifications. Genes Dev. 25, 409–33. doi:10.1101/gad.2021311.critical

Prinz, S., Amon, A., Klein, F., 1997. Isolation of COM1, a new gene required to complete meiotic double- strand break-induced recombination in Saccharomyces cerevisiae. Genetics 146, 781–795.

Putnam, C.D., Hammel, M., Hura, G.L., Tainer, J. a., 2007. X-ray solution scattering (SAXS) combined with crystallography and computation: Defining accurate macromolecular structures, conformations and assemblies in solution. Q. Rev. Biophys. 40, 191–285. doi:10.1017/S0033583507004635

Qvist, P., Huertas, P., Jimeno, S., Nyegaard, M., Hassan, M.J., Jackson, S.P., Børglum, A.D., 2011. CtIP Mutations Cause Seckel and Jawad Syndromes. PLoS Genet. 7, e1002310. doi:10.1371/journal.pgen.1002310

Rajendran, P., Kidane, A.I., Yu, T.W., Dashwood, W.M., Bisson, W.H., Löhr, C. V., Ho, E., Williams, D.E., Dashwood, R.H., 2013. HDAC turnover, CtIP acetylation and dysregulated DNA damage signaling in colon cancer cells treated with sulforaphane and related dietary isothiocyanates. Epigenetics 8. doi:10.4161/epi.24710

Rambo, R.P., Tainer, J. a., 2013. Super-Resolution in Solution X-Ray Scattering and Its Applications to Structural Systems Biology. Annu. Rev. Biophys. 42, 415–441. doi:10.1146/annurev-biophys-083012-130301

Ramírez-Lugo, J.S., Yoo, H.Y., Yoon, S.J., Dunphy, W.G., 2011. CtIP interacts with TopBP1 and Nbs1 in the response to double-stranded DNA breaks (DSBs) in Xenopus egg extracts. Cell Cycle 10, 469–480. doi:10.4161/cc.10.3.14711

Reczek, C.R., Shakya, R., Miteva, Y., Szabolcs, M., Ludwig, T., Baer, R., 2016. The DNA resection protein CtIP promotes mammary tumorigenesis. Oncotarget 7, 32172–83. doi:10.18632/oncotarget.8605

Reinhardt, H.C., Aslanian, A.S., Lees, J. a., Yaffe, M.B., 2007. p53-Deficient Cells Rely on ATM- and ATR-Mediated Checkpoint Signaling through the p38MAPK/MK2 Pathway for Survival after DNA Damage. Cancer Cell 11, 175–189. doi:10.1016/j.ccr.2006.11.024

Reinhardt, H.C., Schumacher, B., 2012. The p53 network: cellular and systemic DNA damage responses in aging and cancer. Trends Genet. 28, 128–136. doi:10.1016/j.tig.2011.12.002

Ribes-Zamora, A., Mihalek, I., Lichtarge, O., Bertuch, A. a., 2007. Distinct faces of the Ku heterodimer mediate DNA repair and telomeric functions. Nat. Struct. Mol. Biol. 14, 301–307. doi:10.1038/nsmb1214

Rivera-Calzada, A., Spagnolo, L., Pearl, L.H., Llorca, O., 2007. Structural model of full-length human Ku70-Ku80 heterodimer and its recognition of DNA and DNA-PKcs. EMBO Rep. 8, 56–62. doi:10.1038/sj.embor.7400847

Rossi, A.M., Taylor, C.W., 2011. Analysis of protein-ligand interactions by fluorescence polarization. Nat. Protoc. 6, 365–87. doi:10.1038/nprot.2011.305

Roy, R., Chun, J., Powell, S.N., 2011. BRCA1 and BRCA2: different roles in a common pathway of genome protection. Nat. Rev. Cancer 12, 68–78. doi:10.1038/nrc3181

Rupert, B.Y.C.S., Dgal, S.O.L.H.G.X., Herriott, R.M., 1958. PHOTOREACTIVATION IN VITRO OF ULTRAVIOLET INACTIVATED HEMOPHILUS INFLUENZAE TRANSFORMING FACTOR 41.

Rzechorzek, N.J., Blackwood, J.K., Bray, S.M., Maman, J.D., Pellegrini, L., Robinson, N.P., 2014. Structure of the hexameric HerA ATPase reveals a mechanism of translocation-coupled DNA-end processing in archaea. Nat. Commun. 5. doi:10.1038/ncomms6506

Säbel, C.E., Shepherd, J.L., Siemann, S., 2009. A direct spectrophotometric method for the simultaneous determination of zinc and cobalt in metalloproteins using 4-(2-

pyridylazo)resorcinol. Anal. Biochem. 391, 74–76. doi:10.1016/j.ab.2009.05.007
Sale, J.E., 2013. Translesion DNA synthesis and mutagenesis in eukaryotes. Cold Spring Harb. Perspect. Biol. doi:10.1101/cshperspect.a012708
Sale, J.E., Lehmann, A.R., Woodgate, R., 2012. Y-family DNA polymerases and their role in tolerance of cellular DNA damage. Nat. Rev. Mol. Cell Biol. 13, 141–152. doi:10.1038/nrm3289
Salesse, S., Verfaillie, C.M., 2002. BCR/ABL: From molecular mechanisms of leukemia induction to treatment of chronic myelogenous leukemia. Oncogene 21, 8547–8559. doi:10.1038/sj.onc.1206082
Salvador, J.M., Brown-Clay, J.D., Fornace, A.J.J., 2013. Gadd45 in stress signaling, cell cycle control, and apoptosis. Adv. Exp. Med. Biol. 793, 1–19. doi:10.1007/978-1-4614-8289-5_1
San Filippo, J., Sung, P., Klein, H., 2008. Mechanism of Eukaryotic Homologous Recombination. Annu. Rev. Biochem. 77, 229–257. doi:10.1146/annurev.biochem.77.061306.125255
Sancar, A., Lindsey-Boltz, L.A., Ünsal-Kaçmaz, K., Linn, S., 2004. Molecular Mechanisms of Mammalian DNA Repair and the DNA Damage Checkpoints. Annu. Rev. Biochem. 73, 39–85. doi:10.1146/annurev.biochem.73.011303.073723
Sarangi, P., Steinacher, R., Altmannova, V., Fu, Q., Paull, T.T., Krejci, L., Whitby, M.C., Zhao, X., 2015. Sumoylation Influences DNA Break Repair Partly by Increasing the Solubility of a Conserved End Resection Protein. PLoS Genet. 11, e1004899. doi:10.1371/journal.pgen.1004899
Sartori, A. a, Lukas, C., Coates, J., Mistrik, M., Fu, S., Bartek, J., Baer, R., Lukas, J., Jackson, S.P., 2007. Human CtIP promotes DNA end resection. Nature 450, 509–14. doi:10.1038/nature06337
Saul, R.L., Ames, B.N., 1986. Background levels of DNA damage in the population 529–530.
Sausville, E.A., 2002. Complexities in the development of cyclin-dependent kinase inhibitor drugs. Trends Mol. Med. 8, S32–S37. doi:10.1016/S1471-4914(02)02308-0
Schaeper, U., Subramanian, T., Lim, L., Boyd, J.M., Chinnadurai, G., 1998. Interaction between a Cellular Protein That Binds to the C-terminal Region of Adenovirus E1A (CtBP) and a Novel Cellular Protein Is Disrupted by E1A through a Conserved PLDLS Motif. J. Biol. Chem. 273, 8549–8552. doi:10.1074/jbc.273.15.8549
Schatz, D.G., Swanson, P.C., 2011. V(D)J Recombination: Mechanisms of Initiation. Annu. Rev. Genet. 45, 167–202. doi:10.1146/annurev-genet-110410-132552
Schipler, A., Iliakis, G., 2013. DNA double-strand-break complexity levels and their possible contributions to the probability for error-prone processing and repair pathway choice. Nucleic Acids Res. 41, 7589–7605. doi:10.1093/nar/gkt556
Schmitt, M.W., Matsumoto, Y., Loeb, L. a., 2009. High fidelity and lesion bypass capability of human DNA polymerase ?? Biochimie 91, 1163–1172. doi:10.1016/j.biochi.2009.06.007
Scott, D.J., Harding, S.E., Rowe, A.J. (Eds.), 2005. A Brief Introduction to the Analytical Ultracentrifugation of Proteins for Beginners, in: Analytical Ultracentrifugation: Techniques and Methods. The Royal Society of Chemistry, pp. 1–25. doi:10.1039/9781847552617-00001
Seol, J.H., Shim, E.Y., Lee, S.E., 2018. Microhomology-mediated end joining: Good, bad and ugly. Mutat. Res. - Fundam. Mol. Mech. Mutagen. 809, 81–87. doi:10.1016/j.mrfmmm.2017.07.002
Setlow, R.B., 1966. Cyclobutane-Type Pyrimidine Dimers in Polynucleotides.
Sfeir, A., Symington, L.S., 2015. Microhomology-Mediated End Joining: A Back-up Survival Mechanism or Dedicated Pathway? Trends Biochem. Sci. 40, 701–714. doi:10.1016/j.tibs.2015.08.006
Shahar, O.D., Ram, E.V.S.R., Shimshoni, E., Hareli, S., Meshorer, E., Goldberg, M., 2012. Live imaging of induced and controlled DNA double-strand break formation reveals extremely low repair by homologous recombination in human cells. Oncogene 31, 3495–3504. doi:10.1038/onc.2011.516
Shaheen, R., Faqeih, E., Ansari, S., Abdel-salam, G., Al-hassnan, Z.N., Al-shidi, T., Alomar, R., Sogaty, S., Alkuraya, F.S., 2014. Genomic analysis of primordial dwarfism reveals novel disease genes 291–299. doi:10.1101/gr.160572.113.
Shaltiel, I. a., Krenning, L., Bruinsma, W., Medema, R.H., 2015. The same, only different - DNA

damage checkpoints and their reversal throughout the cell cycle. J. Cell Sci. 128, 607–620. doi:10.1242/jcs.163766

Shamanna, R.A., Lu, H., De Freitas, J.K., Tian, J., Croteau, D.L., Bohr, V.A., 2016. WRN regulates pathway choice between classical and alternative non-homologous end joining. Nat. Commun. 7, 1–12. doi:10.1038/ncomms13785

Shibata, a., Barton, O., Noon, a. T., Dahm, K., Deckbar, D., Goodarzi, a. a., Lobrich, M., Jeggo, P. a., 2010. Role of ATM and the Damage Response Mediator Proteins 53BP1 and MDC1 in the Maintenance of G2/M Checkpoint Arrest. Mol. Cell. Biol. 30, 3371–3383. doi:10.1128/MCB.01644-09

Shibata, A., Conrad, S., Birraux, J., Geuting, V., Barton, O., Ismail, A., Kakarougkas, A., Meek, K., Taucher-Scholz, G., Löbrich, M., Jeggo, P.A., 2011. Factors determining DNA double-strand break repair pathway choice in G2 phase. EMBO J. 30, 1079–1092. doi:10.1038/emboj.2011.27

Shibata, A., Moiani, D., Arvai, A.S., Perry, J., Harding, S.M., Genois, M.M., Maity, R., van Rossum-Fikkert, S., Kertokalio, A., Romoli, F., Ismail, A., Ismalaj, E., Petricci, E., Neale, M.J., Bristow, R.G., Masson, J.Y., Wyman, C., Jeggo, P. a., Tainer, J. a., 2014. DNA Double-Strand Break Repair Pathway Choice Is Directed by Distinct MRE11 Nuclease Activities. Mol. Cell 53, 7–18. doi:10.1016/j.molcel.2013.11.003

Shiloh, Y., Ziv, Y., 2013. The ATM protein kinase: Regulating the cellular response to genotoxic stress, and more. Nat. Rev. Mol. Cell Biol. 14, 197–210. doi:10.1038/nrm3546

Simsek, D., Brunet, E., Wong, S.Y.W., Katyal, S., Gao, Y., McKinnon, P.J., Lou, J., Zhang, L., Li, J., Rebar, E.J., Gregory, P.D., Holmes, M.C., Jasin, M., 2011. DNA ligase III promotes alternative nonhomologous end-joining during chromosomal translocation formation. PLoS Genet. 7, 1–11. doi:10.1371/journal.pgen.1002080

Sizemore, S.T., Zhang, M., Cho, J.H., Sizemore, G.M., Hurwitz, B., Kaur, B., Lehman, N.L., Ostrowski, M.C., Robe, P. a., Miao, W., Wang, Y., Chakravarti, A., Xia, F., 2018. Pyruvate kinase M2 regulates homologous recombination-mediated DNA double-strand break repair. Cell Res. doi:10.1038/s41422-018-0086-7

Sokka, M., Koalick, D., Hemmerich, P., Syväoja, J.E., Pospiech, H., 2018. The ATR-activation domain of TopBP1 is required for the suppression of origin firing during the S phase. Int. J. Mol. Sci. 19. doi:10.1016/j.ultsonch.2015.02.005

Soria-Bretones, I., Cepeda-García, C., Checa-Rodriguez, C., Heyer, V., Reina-San-Martin, B., Soutoglou, E., Huertas, P., 2017. DNA end resection requires constitutive sumoylation of CtIP by CBX4. Nat. Commun. 8. doi:10.1038/s41467-017-00183-6

Soutoglou, E., Dorn, J.F., Sengupta, K., Jasin, M., Nussenzweig, A., Ried, T., Danuser, G., Misteli, T., 2007. Positional stability of single double-strand breaks in mammalian cells. Nat. Cell Biol. 9, 675–682. doi:10.1038/ncb1591

Spagnolo, L., Rivera-Calzada, A., Pearl, L.H., Llorca, O., 2006. Three-Dimensional Structure of the Human DNA-PKcs/Ku70/Ku80 Complex Assembled on DNA and Its Implications for DNA DSB Repair. Mol. Cell 22, 511–519. doi:10.1016/j.molcel.2006.04.013

Spycher, C., Miller, E.S., Townsend, K., Pavic, L., Morrice, N. a., Janscak, P., Stewart, G.S., Stucki, M., 2008. Constitutive phosphorylation of MDC1 physically links the MRE11-RAD50-NBS1 complex to damaged chromatin. J. Cell Biol. 181, 227–240. doi:10.1083/jcb.200709008

Steger, M., Murina, O., Hühn, D., Ferretti, L.P., Walser, R., Hänggi, K., Lafranchi, L., Neugebauer, C., Paliwal, S., Janscak, P., Gerrits, B., Del Sal, G., Zerbe, O., Sartori, A. a, 2013. Prolyl isomerase PIN1 regulates DNA double-strand break repair by counteracting DNA end resection. Mol. Cell 50, 333–43. doi:10.1016/j.molcel.2013.03.023

Steinke, V., Engel, C., Büttner, R., Schackert, H.K., Schmiegel, W.H., Propping, P., 2013. Hereditary Nonpolyposis Colorectal Cancer (HNPCC)/Lynch Syndrome. Dtsch. Arztebl. Int. 110, 32–38. doi:10.3238/arztebl.2013.0032

Stevens, C., Smith, L., La Thangue, N.B., 2003. Chk2 activates E2F-1 in response to DNA damage. Nat. Cell Biol. 5, 401–409. doi:10.1038/ncb974

Stokes, P.H., Liew, C.W., Kwan, a H., Foo, P., Barker, H.E., Djamirze, a, O'Reilly, V., Visvader, J.E., Mackay, J.P., Matthews, J.M., 2013. Structural basis of the interaction of the breast cancer oncogene LMO4 with the tumour suppressor CtIP/RBBP8. J. Mol. Biol. 425, 1101–10. doi:10.1016/j.jmb.2013.01.017

Stokes, P.H., Thompson, L.S., Marianayagam, N.J., Matthews, J.M., 2007. Dimerization of CtIP may stabilize in vivo interactions with the Retinoblastoma-pocket domain. Biochem. Biophys. Res. Commun. 354, 197–202. doi:10.1016/j.bbrc.2006.12.178

Stucki, M., Jackson, S.P., 2006. γH2AX and MDC1: Anchoring the DNA-damage-response machinery to broken chromosomes. DNA Repair (Amst). 5, 534–543. doi:10.1016/j.dnarep.2006.01.012

Sugasawa, K., Akagi, J.I., Nishi, R., Iwai, S., Hanaoka, F., 2009. Two-Step Recognition of DNA Damage for Mammalian Nucleotide Excision Repair: Directional Binding of the XPC Complex and DNA Strand Scanning. Mol. Cell 36, 642–653. doi:10.1016/j.molcel.2009.09.035

Sugasawa, K., Masutani, C., Uchida, a, Maekawa, T., van der Spek, P.J., Bootsma, D., Hoeijmakers, J.H., Hanaoka, F., 1996. HHR23B, a human Rad23 homolog, stimulates XPC protein in nucleotide excision repair in vitro. Mol. Cell. Biol. 16, 4852–61. doi:10.1128/mcb.16.9.4852

Sum, E.Y.M., Peng, B., Yu, X., Chen, J., Byrne, J., Lindeman, G.J., Visvader, J.E., 2002. The LIM domain protein LMO4 interacts with the cofactor CtIP and the tumor suppressor BRCA1 and inhibits BRCA1 activity. J. Biol. Chem. 277, 7849–7856. doi:10.1074/jbc.M110603200

Sun, M., 2013. Structural and functional characterisation of Human DNA repair protein CtIP/RBBP8. University of Cambridge.

Svergun, D.I., Koch, M.H.J., 2003. Small-angle scattering studies of biological macromolecules in solution. Reports Prog. Phys. 66, 1735–1782. doi:10.1088/0034-4885/66/10/R05

Svilar, D., Goellner, E.M., Almeida, K.H., Sobol, R.W., 2011. Base Excision Repair and Lesion-Dependent Subpathways for Repair of Oxidative DNA Damage. Antioxid. Redox Signal. 14, 2491–2507. doi:10.1089/ars.2010.3466

Szostak, J.W., Orr-Weaver, T.L., Rothstein, R.J., Stahl, F.W., 1983. The double-strand-break repair model for recombination. Cell 33, 25–35. doi:10.1016/0092-8674(83)90331-8

Thangaraju, M., Kaufmann, S.H., Couch, F.J., 2000. BRCA1 facilitates stress-induced apoptosis in breast and ovarian cancer cell lines. J. Biol. Chem. 275, 33487–33496. doi:10.1074/jbc.M005824200

Thornby Bak, S., Sakellariou, D., Peña-Diaz, J., 2014. The dual nature of mismatch repair as antimutator and mutator: For better or for worse, Frontiers in genetics. doi:10.3389/fgene.2014.00287

Ting, N.S.Y., Lee, W.H., 2004. The DNA double-strand break response pathway: Becoming more BRCAish than ever. DNA Repair (Amst). 3, 935–944. doi:10.1016/j.dnarep.2004.03.026

Tripathi, V., Agarwal, H., Priya, S., Batra, H., Modi, P., Pandey, M., Saha, D., Raghavan, S.C., Sengupta, S., 2018. MRN complex-dependent recruitment of ubiquitylated BLM helicase to DSBs negatively regulates DNA repair pathways. Nat. Commun. 9, 1–13. doi:10.1038/s41467-018-03393-8

Truong, L.N., Li, Y., Shi, L.Z., Hwang, P.Y.-H., He, J., Wang, H., Razavian, N., Berns, M.W., Wu, X., 2013. Microhomology-mediated End Joining and Homologous Recombination share the initial end resection step to repair DNA double-strand breaks in mammalian cells. Proc. Natl. Acad. Sci. 110, 7720 LP-7725.

Tyson, J.J., Novak, B., 2008. Temporal Organization of the Cell Cycle. Curr. Biol. 18, R759–R768. doi:10.1016/j.cub.2008.07.001

Uanschou, C., Siwiec, T., Pedrosa-Harand, A., Kerzendorfer, C., Sanchez-Moran, E., Novatchkova, M., Akimcheva, S., Woglar, A., Klein, F., Schlögelhofer, P., 2007. A novel plant gene essential for meiosis is related to the human CtIP and the yeast COM1/SAE2 gene. EMBO J. 26, 5061–5070. doi:10.1038/sj.emboj.7601913

Unno, J., Itaya, A., Taoka, M., Sato, K., Tomida, J., Sakai, W., Sugasawa, K., Ishiai, M., Ikura, T., Isobe, T., Kurumizaka, H., Takata, M., 2014. FANCD2 binds CtIP and regulates DNA-end resection

during DNA interstrand crosslink repair. Cell Rep. 7, 1039–47. doi:10.1016/j.celrep.2014.04.005
Valko, M., Rhodes, C.J., Moncol, J., Izakovic, M., Mazur, M., 2006. Free radicals, metals and antioxidants in oxidative stress-induced cancer. Chem. Biol. Interact. 160, 1–40. doi:10.1016/j.cbi.2005.12.009
Visconti, R., Della Monica, R., Grieco, D., 2016. Cell cycle checkpoint in cancer: A therapeutically targetable double-edged sword. J. Exp. Clin. Cancer Res. 35, 1–8. doi:10.1186/s13046-016-0433-9
Wade Harper, J., Adami, G.R., Wei, N., Keyomarsi, K., Elledge, S.J., 1993. The p21 Cdk-interacting protein Cip1 is a potent inhibitor of G1 cyclin-dependent kinases. Cell 75, 805–816. doi:10.1016/0092-8674(93)90499-G
Wajant, H., 2002. The Fas Signaling Pathway: More Than a Paradigm. Science (80-.). 296, 1635 LP-1636.
Walker, J.R., Corpina, R. a., Goldberg, J., 2001. Structure of the Ku heterodimer bound to dna and its implications for double-strand break repair. Nature 412, 607–614. doi:10.1038/35088000
Wang, H., Li, Y., Truong, L.N., Shi, L.Z., Hwang, P.Y.-H., He, J., Do, J., Cho, M.J., Li, H., Negrete, A., Shiloach, J., Berns, M.W., Shen, B., Chen, L., Wu, X., 2014. CtIP maintains stability at common fragile sites and inverted repeats by end resection-independent endonuclease activity. Mol. Cell 54, 1012–21. doi:10.1016/j.molcel.2014.04.012
Wang, H., Qiu, Z., Liu, B., Wu, Y., Ren, J., Liu, Y., Zhao, Y., Wang, Y., Hao, S., Li, Z., Peng, B., Xu, X., 2018. PLK1 targets CtIP to promote microhomology-mediated end joining. Nucleic Acids Res. 1–16. doi:10.1093/nar/gky810
Wang, H., Shao, Z., Shi, L.Z., Hwang, P.Y.H., Truong, L.N., Berns, M.W., Chen, D.J., Wu, X., 2012. CtIP protein dimerization is critical for its recruitment to chromosomal DNA double-stranded breaks. J. Biol. Chem. 287, 21471–21480. doi:10.1074/jbc.M112.355354
Wang, H., Shi, L.Z., Wong, C.C.L., Han, X., Hwang, P.Y.-H., Truong, L.N., Zhu, Q., Shao, Z., Chen, D.J., Berns, M.W., Yates, J.R., Chen, L., Wu, X., 2013. The interaction of CtIP and Nbs1 connects CDK and ATM to regulate HR-mediated double-strand break repair. PLoS Genet. 9, e1003277. doi:10.1371/journal.pgen.1003277
Wang, H., Xu, X., 2017. Microhomology-mediated end joining: New players join the team. Cell Biosci. 7, 4–9. doi:10.1186/s13578-017-0136-8
Wang, S., El-Deiry, W.S., 2003. TRAIL and apoptosis induction by TNF-family death receptors. Oncogene 22, 8628.
Wang, W., Daley, J.M., Kwon, Y., Xue, X., Krasner, D.S., Miller, A.S., Nguyen, K.A., Williamson, E.A., Shim, E.Y., Lee, S.E., Hromas, R., Sung, P., 2018. A DNA nick at Ku-blocked double-strand break ends serves as an entry site for exonuclease 1 (Exo1) or Sgs1-Dna2 in long-range DNA end resection. J. Biol. Chem. 1, jbc.RA118.004769. doi:10.1074/jbc.RA118.004769
Ward, I.M., Minn, K., Chen, J., 2004. UV-induced Ataxia-telangiectasia-mutated and Rad3-related (ATR) Activation Requires Replication Stress. J. Biol. Chem. 279, 9677–9680. doi:10.1074/jbc.C300554200
Ward, J.F., Evans, J.W., Limoli, C.L., 1987. Radiation and hydrogen peroxide induced free radical damage to DNA. 105–109.
Waterhouse, A.M., Procter, J.B., Martin, D.M. a., Clamp, M., Barton, G.J., 2009. Jalview Version 2-A multiple sequence alignment editor and analysis workbench. Bioinformatics 25, 1189–1191. doi:10.1093/bioinformatics/btp033
Weber, A.M., Ryan, A.J., 2015. ATM and ATR as therapeutic targets in cancer. Pharmacol. Ther. 149, 124–138. doi:10.1016/j.pharmthera.2014.12.001
Wiley, john., 1998. Ammonium sulphate precipitation [WWW Document]. Curr. potocols protein Sci.
Williams, G.J., Williams, R.S., Williams, J.S., Moncalian, G., Arvai, A.S., Limbo, O., Guenther, G., Sildas, S., Hammel, M., Russell, P., Tainer, J. a., 2011. ABC ATPase signature helices in Rad50 link nucleotide state to Mre11 interface for DNA repair. Nat. Struct. Mol. Biol. 18, 423–432. doi:10.1038/nsmb.2038

Williams, R.S., Dodson, G.E., Limbo, O., Yamada, Y., Williams, J.S., Guenther, G., Classen, S., Glover, J.N.M., Iwasaki, H., Russell, P., Tainer, J. a, 2009. Nbs1 flexibly tethers Ctp1 and Mre11-Rad50 to coordinate DNA double-strand break processing and repair. Cell 139, 87–99. doi:10.1016/j.cell.2009.07.033

Williams, R.S., Moncalian, G., Williams, J.S., Yamada, Y., Limbo, O., Shin, D.S., Groocock, L.M., Cahill, D., Hitomi, C., Guenther, G., Moiani, D., Carney, J.P., Russell, P., Tainer, J. a., 2008. Mre11 Dimers Coordinate DNA End Bridging and Nuclease Processing in Double-Strand-Break Repair. Cell 135, 97–109. doi:10.1016/j.cell.2008.08.017

Wong, a K., Ormonde, P. a, Pero, R., Chen, Y., Lian, L., Salada, G., Berry, S., Lawrence, Q., Dayananth, P., Ha, P., Tavtigian, S. V, Teng, D.H., Bartel, P.L., 1998. Characterization of a carboxy-terminal BRCA1 interacting protein. Oncogene 17, 2279–2285. doi:10.1038/sj.onc.1202150

Wood, C.W., Woolfson, D.N., 2018. CCBuilder 2.0: Powerful and accessible coiled-coil modeling. Protein Sci. 27, 103–111. doi:10.1002/pro.3279

Wu, G., Lee, W.H., 2006. CtIP, a multivalent adaptor connecting transcriptional regulation, checkpoint control and tumor suppression. Cell Cycle 5, 1592–1596. doi:10.4161/cc.5.15.3127

Wulhfard, S., Baldi, L., Hacker, D.L., Wurm, F., 2010. Valproic acid enhances recombinant mRNA and protein levels in transiently transfected Chinese hamster ovary cells. J. Biotechnol. 148, 128–132. doi:10.1016/j.jbiotec.2010.05.003

Wyatt, P.J., 1993. Light scattering and the absolute characterization of macromolecules. Anal. Chim. Acta 272, 1–40. doi:10.1016/0003-2670(93)80373-S

Xu, G., Chapman, J.R., Brandsma, I., Yuan, J., Mistrik, M., Bouwman, P., Bartkova, J., Gogola, E., Warmerdam, D., Barazas, M., Jaspers, J.E., Watanabe, K., Pieterse, M., Kersbergen, A., Sol, W., Celie, P.H.N., Schouten, P.C., van den Broek, B., Salman, A., Nieuwland, M., de Rink, I., de Ronde, J., Jalink, K., Boulton, S.J., Chen, J., van Gent, D.C., Bartek, J., Jonkers, J., Borst, P., Rottenberg, S., 2015. REV7 counteracts DNA double-strand break resection and affects PARP inhibition. Nature. doi:10.1038/nature14328

Yamamoto, A., Taki, T., Yagi, H., Habu, T., Yoshida, K., Yoshimura, Y., Yamamoto, K., Matsushiro, A., Nishimune, Y., Morita, T., 1996. Cell cycle-dependent expression of the mouse Rad51 gene in proliferating cells. Mol. Gen. Genet. 251, 1–12.

Yamamoto, K., Ishihama, A., 2005. Transcriptional response of Escherichia coli to external zinc. J. Bacteriol. 187, 6333–40. doi:10.1128/JB.187.18.6333-6340.2005

Yang, S., Kuo, C., Bisi, J.E., Kim, M.K., 2002. PML-dependent apoptosis after DNA damage is regulated by the checkpoint kinase hCds1/Chk2. Nat. Cell Biol. 4, 865–870. doi:10.1038/ncb869

Yeo, J.E., Lee, E.H., Hendrickson, E. a, Sobeck, A., 2014. CtIP mediates replication fork recovery in a FANCD2-regulated manner. Hum. Mol. Genet. 23, 3695–705. doi:10.1093/hmg/ddu078

Yoshizawa-Sugata, N., Masai, H., 2009. Roles of human AND-1 in chromosome transactions in S phase. J. Biol. Chem. 284, 20718–20728. doi:10.1074/jbc.M806711200

You, Z., Bailis, J.M., 2010. DNA damage and decisions: CtIP coordinates DNA repair and cell cycle checkpoints. Trends Cell Biol. 20, 402–409. doi:10.1016/j.tcb.2010.04.002

You, Z., Shi, L.Z., Zhu, Q., Wu, P., Zhang, Y.-W., Basilio, A., Tonnu, N., Verma, I.M., Berns, M.W., Hunter, T., 2009. CtIP links DNA double-strand break sensing to resection. Mol. Cell 36, 954–69. doi:10.1016/j.molcel.2009.12.002

Yousefzadeh, M.J., Wyatt, D.W., Takata, K.I., Mu, Y., Hensley, S.C., Tomida, J., Bylund, G.O., Doublié, S., Johansson, E., Ramsden, D. a., McBride, K.M., Wood, R.D., 2014. Mechanism of Suppression of Chromosomal Instability by DNA Polymerase POLQ. PLoS Genet. 10. doi:10.1371/journal.pgen.1004654

Yu, X., Baer, R., 2000. Nuclear localization and cell cycle-specific expression of CtIP, a protein that associates with the BRCA1 tumor suppressor. J. Biol. Chem. 275, 18541–18549. doi:10.1074/jbc.M909494199

Yu, X., Fu, S., Lai, M., Baer, R., Chen, J., Breast, B., Susceptibility, C., 2006. BRCA1 ubiquitinates its binding partner CtIP 1, 1721–1726. doi:10.1101/gad.1431006.CtIP

Yuan, J., Chen, J., 2009. N terminus of CtIP is critical for homologous recombination-mediated double-strand break repair. J. Biol. Chem. 284, 31746–52. doi:10.1074/jbc.M109.023424
Yun, M.H., Hiom, K., 2009. CtIP-BRCA1 modulates the choice of DNA double-strand-break repair pathway throughout the cell cycle. Nature 459, 460–3. doi:10.1038/nature07955
Zetterberg, A., Larsson, O., Wiman, klas G., 1995. What is the restriction point? curent Opin. cell Biol. 7. doi:10.1161/JAHA.116.003407
Zha, S., Boboila, C., Alt, F.W., 2009. Mre11: Roles in DNA repair beyond homologous recombination. Nat. Struct. Mol. Biol. 16, 798–800. doi:10.1038/nsmb0809-798
Zhan, Q., Antinore, M.J., Wang, X.W., Carrier, F., Smith, M.L., Harris, C.C., Fornace Jr, A.J., 1999. Association with Cdc2 and inhibition of Cdc2/Cyclin B1 kinase activity by the p53-regulated protein Gadd45. Oncogene 18, 2892.
Zhang, F., Tang, H., Jiang, Y., Mao, Z., 2017. The transcription factor GATA3 is required for homologous recombination repair by regulating CtIP expression. Oncogene 36, 5168–5176. doi:10.1038/onc.2017.127
Zhou, Z.W., Liu, C., Li, T.L., Bruhn, C., Krueger, A., Min, W.K., Wang, Z.Q., Carr, A.M., 2013. An Essential Function for the ATR-Activation-Domain (AAD) of TopBP1 in Mouse Development and Cellular Senescence. PLoS Genet. 9. doi:10.1371/journal.pgen.1003702
Zhu, W., Ukomadu, C., Jha, S., Senga, T., Dhar, S.K., Wohlschlegel, J. a., Nutt, L.K., Kornbluth, S., Dutta, A., 2007. Mcm10 and And-1/CTF4 recruit DNA polymerase α to chromatin for initiation of DNA replication. Genes Dev. 21, 2288–2299. doi:10.1101/gad.1585607
Zhu, Z., Chung, W.H., Shim, E.Y., Lee, S.E., Ira, G., 2008. Sgs1 Helicase and Two Nucleases Dna2 and Exo1 Resect DNA Double-Strand Break Ends. Cell 134, 981–994. doi:10.1016/j.cell.2008.08.037
Zimmermann, M., Lottersberger, F., Buonomo, S.B., Sfeir, A., de Lange, T., 2013. 53BP1 Regulates DSB Repair Using Rif1 to Control 5′ End Resection. Science (80-.). 339, 700 LP-704.
Zou, L., Elledge, S.J., 2003. Sensing DNA Damage Through ATRIP Recognition of RPA-ssDNA Complexes. Science (80-.). 300, 1542 LP-1548.